Towards a unified model of Pavlovian conditioning: Solution to the extinction problem

KRYUKOV V.I. (HEGUMEN THEOPHAN)
St. Daniel Monastery,
Danilovsky Val 22 Moscow, 115191,
RUSSIA
kryukov@msdm.ru

Abstract: - We are proposing an oscillatory habituation mechanism of extinction that can integrate and reconcile some computational and behavior mechanisms in the unified model of Pavlovin conditioning. The suggested unified model explains most of the recent data on extinction, including renewal, reinstatement, reacquisition, and spontaneous recovery and predicts that extinction can incorporate new learning with some unlearning. This model is the same as the "Neurolocator" model [23] with some minor modifications to account for timing and trace conditioning. The main characteristics of this model are as follows: oscillatory binding of CS-US representation; non-Hebbian learning by synchrony; novelty guided attention; parallel character of interaction between all brain structures through the septal pacemaker, and global control from the septo-hippocampal system. This model offers viable answers to a number of questions on the extinction problem, posed in current literature. and suggests a new neurocognitive mechanism of anxiety.

Key-Words: - Extinction, habituation, attention, renewal, reinstatement, reacquisition, spontaneous recovery, anxiety

1 Introduction

Extinction is a complex phenomenon that has resisted simple explanation in terms of computational models [35]. It occurs when after the CS-US association learning, a CS is presented alone, without the US, for a number of trials and eventually the conditioned reflex (CR) is diminished or eliminated. Although the response to the conditioned stimulus is attenuated during the extinction, the original association is surprisingly unaffected. Pavlov's [39] own investigations showed that extinction does not erase the original excitatory learning. Since then, the number of candidates for down-regulating responding while preserving the original learning has continued to grow (for the recent reviews see [28], [42], [14], [45], [36], [44]. That fact testify that we are far from a clear understanding of extinction nature.

To cope with the ever growing mass of data and theories on extinction we propose to use an oscillatory mechanism as a conceptual limiting condition for new models of extinction. For example, the synchrony of hippocampus and amygdala in theta frequency range increases during consolidation [52] and reconsolidation [39] of fear memories, while theta synchrony decreases at remote memory stages [39], and during fear memory extinction [41]. In addition, theta phase relations between regions vary characteristically during different states of fear memory [41]. This speaks for the fact that theta activity obviously reflects some fundamental mechanisms of conditioning and extinction. None of the existing models or theories can account for these facts.

Another mechanism that seems to be useful in search of a unified model of conditioning, is a well-known phenomenon of habituation. From the previous attempts to build a unified view on extinction problem we consider that the work of [33] is the most fruitful. They suggested that many of the characteristics of behavior undergoing extinction may result from a simpler process of habituation. In particular, behavior undergoing extinction shows 12 of the fundamental properties of behavior undergoing habituation. Kamprath et al [16] underscore the importance of habituation as a determinant of fear extinction.

And finally, the third mechanism for the unified model of extinction is proposed by developmental studies of fear extinction in rats (e.g. [18]) that helps to reveal specific "unlearning" neuronal circuits implicated in the failure to acquire or maintain extinction memories. The current interest in this field has motivated researchers to develop new therapeutic strategies for human anxiety disorders and related psychiatric conditions [14]. But we consider the developmental studies as another limiting condition to reject many models and theories that fail to integrate this "unlearning" mechanism with a currently dominant "new learning" mechanism in the interests of general theory of memory.

Combining these tree mechanisms we are proposing a unified model for the solution of the extinction problem which is more easily appreciated with a brief review of previous behavioral and computational extinction mechanisms: *unlearning, new learning, habituation, and multiple mechanisms.*

Unlearning: It is perhaps the simplest associative mechanism of extinction in which the excitatory association between the CS and US representations formed during acquisition is weakened and ultimately broken through extinction training. For example, extinction in cerebellar learning is considered to be a better example of unlearning [31]. In particular, inhibition of climbing fibers of the cerebellum in the model of eye blink conditioning leads to “unlearning” of conditioned blinking response [20]. However, unlearning currently is considered by most investigators to be untenable because it does not easily account for observations of CR recovery following extinction [3], [36] although according to [8] some unlearning does take place during extinction.

New learning: It is Pavlov's original notion that extinction is new learning, rather than erasure of conditioning in which firstly, the excitatory association emerges from extinction training relatively intact and secondly, inhibitory association forms which effect upon the US representation is opposite to that of the excitatory association. This “inhibitory” mechanism is common in the connectionist models of extinction, which describes extinction as a generation and strengthening of a second, inhibitory association between the CS and US representations, which act concurrently with the excitatory association and directly opposes the tendency of the excitatory association to activate the US representation. The simulations results suggest that extinction in fear conditioning is more akin to a new learning [31]. As a whole, according [35] the search for an inhibitory brain structure has not been very fruitful, as there has not been worked out one structure which putative role in extinction has not been met with substantial empirical challenges.

Habituation: This hypothesis seems to challenge the above mentioned “inhibitory” associative mechanism, since it was demonstrated that the decrease in conditional response on repeated nonreinforced stimulus presentation following conditioning, shows fundamental properties of habituation as non-associative forms of learning [17]. Moreover, nonassociative mechanisms are using a number of commonalities between extinction and habituation, and these mechanisms are used as a ground for the argument that the response decrement in both cases may arise at least partly through the same mechanisms. Their findings imply that the success of exposure therapies is, at least partially, based on successful habituation. Despite its strength, the habituation hypothesis does not provide a complete explanation which is considered to be determined by multiple processes [33], [35].

Multiple mechanisms of extinction include both associative and nonassociative ones although the exact nature of those mechanisms and the manner in which they interact is not fully understood. For example, unlearning and new learning mechanisms may coexist in animals extinguished 10 min after acquisition exhibited no recovery in any of them, whereas animals extinguished 72 h after acquisition exhibited robust recovery in all cases. This means that different neural mechanisms are recruited in learning, depending on the temporal delay of fear extinction [37]. Another example of multiple mechanisms of extinction is given in the model by Redish et al [46] which accommodates extinction and renewal through two simple processes: (a) the generalized associative learning-by-error mechanism of the Rescorla-Wagner [47] model (that successfully captures the slow extinction, but is unable to capture the quick “relearning” that is renewal) and (b) the situation recognition process that categorizes the observed cues into situations (a kind of familiarity/novelty detector), which can rapidly reinstate original the CS-US association when an animal returns into training context. A careful analysis of this model [9] shows that it cannot fully explain the general context dependent renewal in a context different from the training one. In connection with this model as well as with any previous ones a number of questions arise, that are being considered in our next Section.

2 Problem Formulation

Our understanding of how extinction is best conceptualized has remained rather limited [8]. The current understanding of the neural basis of fear extinction is quite insufficient, compared to the acquisition of the conditioned fear [19]. First of all, the question arises if there is a common neural mechanism underlying the behavior properties of extinction, as described on the left hand side of Table 1. As a tentative answer to this question we propose to identify extinction, with a system habituation which neurocognitive properties are presented on the right hand side of Table 1. Of course, these properties in their turn need explanations in terms of neural processes, and these explanation are presented in our work [23]. Now we should present them to be in the framework of the Pavlovian conditioning.

The main extinction problem is to find a unified computational model that can successfully answer the following questions [44]:

Table 1 - **Main properties of extinction and habituation**

Behavior properties of extinction	Neurocognitive properties of habituation
1. *New learning*. Extinction is the development of a new memory of CS-noUS association that competes with the initial memory of CS-US association for the control of behavior (Pavlov, 1927).	1. *New learning*. Gradiul habituation should be regarded as "negative learning" indicating to parallel formation of corresponding memory trace (Vinogradova, 2001).
2.*Spontaneous recovery*. Expression of extinction decays with time (Robbins, 1990).	2. *Spontaneous recovery* is a fundamental characteristic of habituated responses (McSweeney and Swindell, 2002).
3. *Disinhibition.* Extinguished responding may be restored by presenting a novel, extraneous stimulus (McSweeney and Swindell, 2002)	3. *Dishabituation.* The habituated responses are recovered with any changes in the parameters of signal (Vinogradova, 2001).
4. *Renewal*. Extinction is not expressed as strongly if testing occurs in a context different from training context (Bouton, 2002).	4. *Renewal*. The release from habituation often is taken as an explanation for renewal (McSweeney and Swindell, 2002).
5. *Reinstatement.* If unsignaled USs are presented after extinction with the extinction training context, it causes CR (Bouton, 2002).	5. *Reinstatement.* A partial explanation for reinstatement is sensitization by stimuli from another modality (McSweeney and Swindell, 2002).
6. *Specificity*. The impact of extinction training is relatively specific to the extinguished CS (Herry et al, 2008).	6. *Specificity*. When a stimulus is repeated, responses will persist for a long time in the core system of the stimulus (Yamaguchi, 2004).
7. *Attention*. The attention can sometimes play a role in extinction (Delamater, 2004). The extinction is best construed in terms of attentional decrements (Robbins, 1990).	7. *Attention*. The theta regulated attention is provides most complete explanation for habituation in hippocampus (Vinogradova, 2001).

1) What are the neural circuits of extinction learning and how do they interact with the circuits mediating conditioning?

2) What is the role of hippocampus, prefrontal cortex and amygdala in extinction?

3) Where is (the seat of extinction) plasticity *necessary* for the acquisition of extinction?

3 Problem Solution

Our model belongs to the class of physiologically motivated attentional models, and therefore can in principle explain all of the extinction properties listed in Table 1. Despite the fact that the model originally was proposed for solving long term memory problems, it can easily be adapted to the Pavlovian conditioning as a particular case model of episodic memory. Indeed, here again attention is the key to all effects. It is closely connected with the theta/gamma partial synchronization of the basic brain structures, with a specific function to bind oscillatory representations of CS, US and reactions, so that a CR is possible without a US as a result of learning and partial synchronization. Such synchronization is most easily realized by the introduction of a central oscillator with variable frequency that is acting as a global pacemaker. The simplified star-like architecture with a central oscillator (CO) and peripheral oscillators (POs) is given in Fig.1. Some POs represent CS, some US, and some others represent a final reaction. The association of CS and US through the synchronization in conditioning usually requires some learning because oscillatory representations of CS and US have different non-overlapping theta frequency bands. But repeated presentation of CS, recruiting of new POs and re-circulating activity between CO and POs lead to a closer CS frequency representation until the synchronization of CS and US is possible despite the initial detuning. The flexible control and adaptivity is due to the forward-backward connections of POs

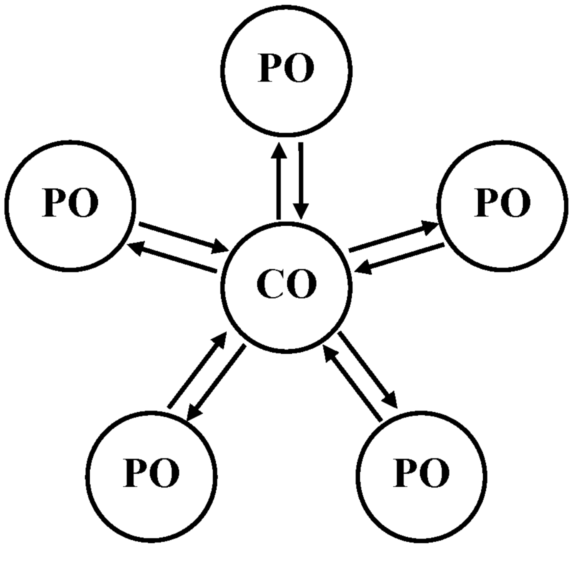

Fig.1. Simplified architecture with central oscillator CO, and peripheral oscillators POs

with CO that can change the current frequency of ensemble synchronization, involving cortical, cerebellar, and amygdalar POs with different natural frequencies in various multimodal ensembles. Accordingly, attention is switched (automatically as well as voluntary) from one group of oscillators to another through this changing frequency of the CO, thus realizing different configural and contextual acquisition, retrieval and extinction. A detailed descriptions of this model, its structure, working principles, and predictions are given elsewhere [23]. Here we are restating in short way some details needed for fear conditioning and extinction model understanding.

The model works like PLL[1] system, which is well known in communication engineering. It comprises five standard modules: a receiver, a voltage controlled oscillator; a phase detector, a low-pass filter, and a summator. The neural representation of this model is sketched in Fig.2, and is based on well established functions of the various parts of the limbic system [57]: the medial septum (MS) is a central pacemaker and a voltage controlled oscillator; the CA3 field of hippocampus is a comparator or a phase detector; the hippocampal fascia dentata (FD) is an input mixer and a receiver of specific inputs; the lateral septum (LS) is an output mixer and a summator of individual lamellas of CA3 fields, i.e., concurrently operating sections of the hippocampal formation are almost independent from each other structurally and functionally [60]. The similar lamellar structure of CA1 field and corresponding parallel pathways of the limbic system is a morphological basis for the tapped delay-line with parallel sub-lines. All these structures, according to [57] are interconnected and form two closed loops, as shown in Fig.3. The first loop deals with information, and includes the hippocampal field CA1, anterior thalamus, neocortex, and other structures which retain, even if partially, their signal-specific sensitivity. This loop is active during the initial information memory formation in neocortex, as well as during the online information treatment performing; for example, in long delays in recycling of signals for the working memory and trace conditioning. The second CA3-based loop, serving for regulating purposes, is responsible for a non-specific brain activation (arousal) and regulation of activating reticular formation. At the same time, the second loop serves the function of a negative feedback for the regulation of the septal oscillator theta frequency, with CA3 being a phase detector or a comparator. As a result, the whole ensemble of POs will be synchronized on the system theta rhythm which is defined by the summary activity of all POs, with the relative salience of corresponding stimuli being taken into account.

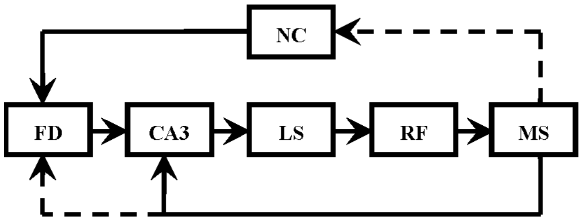

Fig.2 Schematic diagram of the theta-regulated attention model. Abbreviations as in Figure 3

The learning rule is non-Hebbian, as it is based on the following *Isolability Assumption*: when the number of POs oscillators locked in an ensemble reaches a critical value, their physiological labilities tend to be equalized, i.e., the oscillators that are gradually brought to a common rhythm in an ensemble will change their natural frequencies towards a common one, thus implementing isolability coding of information, which is a form of configural coding. Such learning initially may be very fast, sometimes in one-short, while postlearning fixing of new natural frequencies is rather slow, (hours, even days due to the consolidation) and starts after the initial signal retention and some rest or sleep.

[1] A phase-locked loop (PLL) is an electronic control system that generates a signal of controlled oscillator that is locked to the phase of an input signal. A phase-locked loop circuit responds to both the frequency and the phase of the input signals, automatically raising or lowering the frequency of a controlled oscillator until it is matched to the input in both frequency and phase (Lindsey, 1972).

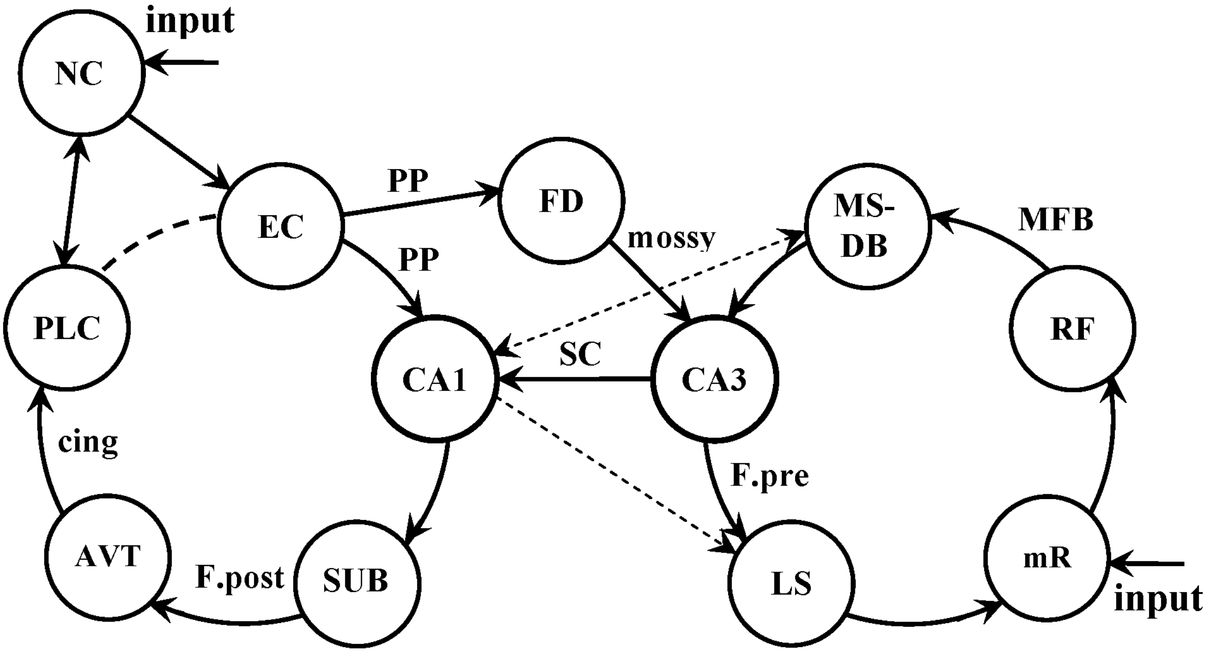

Fig.3 Simplified scheme of two limbic circuits: regulatory and informational (After Vinogradova, 2001). Only principal connections are shown. AVT – antero-ventral nucleus of thalamus; CA1 and CA3 – hippocampal fields; cing – cingulum; FD – fascia dentate; F.pre – fornix precomissuralis; F.post – fornix postcomissuralis; EC –entorhinal cortex; LS – lateral septal nucleus; MFB – medial forebrain bundle; mossy – mossy fibre; MS-DB – medial septal nucleus and nucleus of diagonal band; NC – neocortex; PLC – posterior limbic cortex; PP – perforant path; RF – reticular formation; mR – median raphe nucleus; SC – Schaffer collaterals; SUB – subiculum. New powerful reciprocal connections of CA1 with MS added according to Takacs et al, 2008, as well as well known projections from CA1 to LS according to Risold and Swanson (1996)

According to this model, conditioning is a system process with many brain structures interacting through theta synchronization and with the septo-hippocampal system being a global coordinator of various centers. The frequency detuning between the CS and US is decreased by means of two mechanisms: the first, for a small detuning – by the phase regulation in CA3-based regulatory circle, and the second, for a large detuning – by the CA1-based informational circle (see Fig. 3).

The main predictions/explanations that can be derived from the model are as follows. The neocortical columns may act as a permanent repository of traces in fear conditioning. In particular, there is a long-known fact of the Pavlovian laboratory that decortication prevents habituation in animals. The medial septum may serve as a global pacemaker and (jointly with septo-hippocampal system) as a 'core timer' of variable speed and times. The hippocampus functions as a phase comparator (CA3), or as a delay time comparator (CA1), or both, could be affecting a common septal theta pacemaker to change its frequency in an adaptive way. The CA1-based information circuit can provide a controlled long delay through the reverberating trace of CS in the limbic system. Due to a circular or spiral mode of neural reverberation in this subsystem, the location of peak responses in extinction is not changing. Many other behavioral and physiological effects can be explained by means of this model, some of them (e.g. habituation, dishabituation, novelty detection) have been described in the original "Neurolocator" model of LTM and attention [23]. Here we are presenting preliminary answers to the problematic questions that were stated as the main extinction problem (see previous Section).

1) The neural circuits of extinction learning are largely the same as those of conditioning ones because conditioning is a binding CS-US process, while extinction is an unbinding one and a new learning process; and therefore it is built on the same basic principles, brain structures, and learning rules as conditioning. This prediction is in agreement with the neuroimaging data showing overlapping of neural circuits mediating extinction, reversal of conditioning and regulation of fear, but it also reveals nothing about how these circuits interact between each other [51]. Our model further predicts that such interaction occurs mainly through new learning, unlearning, and their combination. In particular, *unlearning* (though in a small scale) follows from the tendency of cortical oscillators to be recruited into a novelty set of cortical oscillators with a different natural frequency, and

thus it is weakening the original conditioning learning. This tendency has more impact for the unconsolidated memory, i.e. with an unfixed new natural frequency, when extinction is conducted immediately after learning as it have been confirmed by Myers et al [37]. As it is explained below, this tendency is much stronger for the species with undeveloped or damaged mPFC . Similarly, extinction as *new learning* follows from a novelty of association of CS and the context when US is not applied. This process leads to a changing of the theta rhythm frequency of extinction, as compared to that of conditioning thus leaving the original CS-US binding almost intact at their original theta frequency, since different theta attractors are nonoverlapping.

2) The role of hippocampus in extinction according to the "Neurolocator" model is the same one as in LTM formation: hippocampus is a comparator, not a memory store, as it has been supposed to be. This role is similar to that of the error extraction mechanism of an associative learning or the time difference mechanism of connectionist models. The role of PFC is to function as a dominant oscillator among other peripheral cortical and subcortical oscillators, and that is to be the most powerful (salient) among all POs connected with the central oscillator. Therefore, PFC can effectively change the current theta frequency according to instructions, plans, decisions or a change of a context, and thus change the set of oscillators actively involved in the theta regulation. This dominance of PFC can eliminate some cortical oscillators from their active control of the system behavior, without using any inhibitory circuits or dedicated brain structures, but simply by shifting the "center of gravity" for overall POs configuration that is determining the current theta frequency. For example, amygdala being a very important component of fear conditioning of biasing attention towards fearful stimuli [8], can be sometime "inhibited" or habituated [5] without introduction of any inhibitory circuits, in other cases (see later) it can play a decisive role in an overall control of learning and unlearning. But it is important to note that the above biasing functions of PFC and amygdala are only feasible in a global system with hippocampus as a leading structure, because system habituation is unachievable without a comparator.

3) The answer to the question of plasticity in extinction is as follows. Like other types of learning, extinction learning occurs in three phases: acquisition, consolidation, and retrieval. Acquisition of extinction is an initial learning that occurs when conditioned responses are declining within an extinction training session as a result of system habituation and theta desynchronization across many POs. This process is followed by a consolidation phase, lasting several hours, in which physiological and molecular processes stabilize long-term memory for extinction by changing natural frequencies of POs that were synchronized during learning. Subsequent to this, presentation of the extinguished CS triggers a retrieval of extinction memory by reviving learned POs configurations that were active in the extinction acquisition. Poor retrieval of extinction is characterized by high levels of conditioned responding to the extinguished CS, reflecting expression of the original conditioning memory. So learning at each of these phases occurs across many brain structures. Theta synchronization appears to be an important organizing principle for creating time windows of fear memory consolidation within extended hippocampal-amygdala-prefrontal cortical networks. While connectionism usually identify a single structure as a locus of extinction memory (mainly in amygdala), in our unified extinction model, like conditioning itself, memory is distributed across a network of structures. Extinction-related plasticity in each separate structure, however, does not serve identical roles. For example, plasticity in the sensory cortex is necessary to learn CS-noUS associations, whereas LTP plasticity in the hippocampus serves to support the function of switch off for some lamellas, corresponding to strongly activated POs for subsequent consolidation [57]. Similarly, plasticity in amygdala, PFC, cerebellar etc is indexing these structures as the ones taking part in a different task for ensuing consolidation and retrieval. Some important particular cases will be presented in the next paragraphs of this article. However, it is impossible to pinpoint a single site of extinction acquisition, because acquisition is distributed across several structures and its final behavioral output results from the combined influences of many regions on each other, not just a single pathway. Therefore, there naturally arise many contradictory results in the literature on extinction that presume "one structure one function", especially in the case of amygdala.

Below we are discussing three important cases of components plasticity and interaction, concerning three highly important questions of extinction.

Why is extinction memory context dependent while conditioning memory is not? The "Neurolocator" model explains this as follows. The hippocampus play critical role in the context dependent conditioning due to its ability to integrate contextual features in its different lamellas along septo-temporal axis. The steady-state theta frequency is dependent on all stimuli impinging on an organism during conditioning as a weighted sum of their saliencies. Since the US representation is usually the most salient one among all of them, it "overshadows" all less salient stimuli so that the resulting theta frequency is determined mainly by the CS-US association. In the process of extinction the role of the US representation is gradually diminishing through the repetition of CS without US until synchronization between CS and US representations is eventually broken. At this moment the context starts to play a theta regulating role because since that moment it is not "overshadowed" by the US representation. As a result the extinction memory becomes context-dependent, in a sense that for extinction memory to be retrieved the CS and the context should be the same as during the extinction acquisition. Otherwise the renewal of conditioned response will appear regardless of the context which will be "overshadowed" again by the US representation.

This explanation is supported by the following data:

- The molecular requirements of extinction are different from those of fear conditioning in one respect: they are crucial at the time of the first CS-no US contingency, in the 1st retrieval test. They play a role only at the time of the initiation of extinction and shortly thereafter [35], [56].
- Hippocampus not only plays a role in contextual encoding and retrieval of fear extinction memories, but also interacts with other brain structures to regulate context-specificity of fear extinction [12].
- All rats that had received hippocampal inactivation before extinction training demonstrated renewed fear, regardless of the context in which the testing took place. This suggests a role for the dorsal hippocampus in both acquiring the extinction memory and encoding the CS-context relationship that yields the context dependence of extinction. Hippocampus is involved in the acquisition, contextual encoding, and context-dependent retrieval of fear extinction [7].

Why are the stimulus-outcome associations preserved throughout the extinction? First of all let us explain by means of the "Neurolocator" model how the original CS-US associations could be erased. This can happen during the extinction in the course of new learning when a considerable number of POs representing the CS-US attractor could be recruited into a new CS-no US attractor and consolidated there alike "generalization". That is how their natural frequencies are changed considerably enough to forget the former association. Hence, to avoid this "generalization", the PLL system must have a capacity to reject unwanted POs, thus realizing "generalization decrement", which is the main cause of context dependency of associative theory [4]. That means that the catching range of PLL must be regulated by a flexible control of the leading PO's salience and its natural frequency. The most powerful PO is mPFC guided by hippocampus and amygdala and therefore the higher the salience of that PO is, the narrower the catching range is; and the stronger the rejection of extraneous POs is, the better the preservation of previous association throughout the extinction is. On the contrary, the weaker the influence of mPFC is the grater the impairment to the extinction memory is.

The above arguments explain the following facts:

- Lesions of mPFC impair recall of extinction under various conditions, and stimulation of mPFC is strengthening extinction memory [43].
- Temporary inactivation of mPFC at the time of extinction training blocks extinction retention the following day in 24-day old rats, but not in 17-day old rats; and their immunohistochemical analyses revealed that extinction in 17-day old rats does not involve mPFC [18], suggesting that extinction in 17-day old rats relies on an inflexible system that does not allow for the expression of a previously learned fear once it has been extinguished. In other words, extinction may be unlearning at this age, at least in a functional sense.
- Functional imaging studies of PTSD patients exhibit hypoactivity in the vmPFC but hyperactivity in the amygdala. A recent study of brain-injured and trauma-exposed combat veterans confirms that amygdala damage reduces the likelihood of developing PTSD. But contrary to the prediction of the dominant inhibition model, vmPFC damage also reduces the likelihood of developing PTSD [21].
- There exist numerous data on extinction that suggest both "new learning" and "erasure" as mechanisms for extinction ([8], [1], [27], [36]).

According to the "Neurolocator" model it is possible to shift between these two mechanisms by an automatic or voluntary control of selectivity of leading PO, presumably mPFC. When extinction occurs early in a rat's development, the balance between the unlearning and new learning processes of extinction is shifted compared to that of an adult rat, and in the latter extinction relies more on unlearning rather than new learning due to the low selectivity of late developing mPFC. Our hypothesis is that in adult species of animals a shift from an active "new learning" to a passive "erase" mechanism is also possible, and the switch between them is realized by the central nucleus of the amygdala which is recently discovered to be a neural switch for active and passive fear [10], acting irrespective of negative valence [29], [54].

Many psychiatrists around the world tend to consider fear disorders as the ones resulting from the syndrome of extinction deficit (for review see [6]). But is the impaired extinction of an acquired fear the core symptom of anxiety disorders, and why is it resistant to existing pharmacotherapy [34]? The "Neurolocator" model provides the answer that extinction is a system type of process that includes a large-scaled network undergoing system habituation (including hippocampus, mPFC, amygdala, and neocortex). The necessary condition for habituation to occur is the integrity of both inputs into the hippocampal CA3 field because in that case the potentiated synapses of cortical input do not respond to sensory stimuli, terminating reactive state of the CA3 neurons [57]. Moreover, the integrity of the whole regulatory CA3-based loop (see Fig. 3) is necessary for the system habituation, including fornix, lateral septum, raphe nucleus and medial septum. That is why impaired extinction may indeed be core symptoms of anxiety that is resistant to present and probably future pharmacotherapy (cf. [53]).

The above explanation is in accord with the following facts:

- Entorhinal cortex plays a role in extinction [2].
- Even brief periods of intense stress can cause impaired fear extinction [15]. In particular, it may be the result of atrophy of apical dendrites of hippocampus CA3 [59].
- Lesions of medial septal cholinergic neurons impair contextual fear extinction while leaving fear conditioning intact [55].

To illustrated the power of our solution to the extinction problem in this article we are briefly explaining the data which Delamater [8] considers as recent discoveries at the behavioral level of extinction: (1) CS–US stimulus associations specific in their sensory content are fully preserved during extinction; in the unified model extinction as new learning occurs in CS-noUS theta-attractor that is non overlapping with CS-US theta attractor. (2) inhibitory stimulus–response associations appear to be learned during the extinction (“new learning”); in our model such responses are formed by some POs that subject to the same learning rule as the one during conditioning without any inhibitory connections. (3) extinction is influenced by the level of activation of the US representation during nonreinforced trials; according to the “Neurolocator” model the US is represented by the most salient POs which predominantly determine theta frequency and speed of learning. (4) decreases in attention can influence conditioned performance during the extinction; extinction learning in our model follows the system habituation resulting in decrease in global synchronization and selective attention due to the loss of novelty. (5) contexts acquire an ability to modulate learning during both conditioning and extinction, because overlapping neural circuits (hippocampal lamellae) are active in both and because of context “overshadowing” by the US representation depends on progress of extinction.

4 Conclusion

The key to understanding of extinction phenomenon that is the hippocampus playing the role of a comparator, this fact was firmly established by Vinogradova [58], [57] and recently confirmed by a number of studies [24]-[26]. This fact allowed us to build a general computational model of memory and attention [22], [23] which is capable of reconciling the major existing theories on the role of hippocampus in the long-term memory and propose a simple solution to several outstanding problems, concerning the neurobiology of memory such as: consolidation and reconsolidation, persistency of long term memory, novelty detection, habituation, long-term potentiation, and multifrequency oscillatory self-organization of the brain.

In the present paper the very same model is applied to solve the extinction problem as a particular case of general memory and attention problem, and find the answers to a number of difficult questions posed in the extinction literature. Among these questions there is one especially pertaining to this Conference namely the problem of anxiety mechanism and how understanding of the fundamental mechanism of memory can help in finding a new strategy of curing anxiety disorders. We are offering a solution only to the first part of the problem but we believe that by means of the same model in the future we shall be able to solve the second part.

What we have done in this respect can be compared with a well-known solution for the anxiety problem made by Gray and McNaughton [11]. They suggested that septo-hippocampal formation is the seat of anxiety in the brain, and that it acts to detect and resolve the situations of conflicts or uncertainty in animals, and thus protect them from a danger. Having detected a conflict, the hippocampal formation in a way acts to resolve this conflict by increasing the levels of attention and arousal, and through the behavioral inhibition of prior, and on-going motor programs. These behavioral responses constitute anxiety and allow the animal to gather more information in order to resolve the conflict before responding appropriately. Thus, Gray and McNaughton suggested that hippocampal system resolves the conflict by increasing the weight given to affectively negative information. In other words, in a normal animal the hippocampal system will act to favour avoidance behavior over approach behavior.

The model that we are suggesting possesses the following common features with that of Gray and McNaughton: (a) hippocampus acts as a comparator[2]; (b) conflict is resolved by an increasing level of attention and arousal and desynchronization of prior on-going motor POs; (c) mPFC acts as an executive in resolving of conflict by increasing its weight given

[2] The idea of the hippocampus as a comparator was admittedly borrowed by them from Vinogradova (1975). This fact has been later confirmed by McNaughton (2006).

to affective negative information. On the other hand, our model is essentially different from that of Gray and McNaughton in the several ways. In our model (a) oscillatory mechanism coordinates all the components of the memory and attention system; (b) habituation in part is a neurocognitive mechanism of conflict resolving; and (c) developing structures of mPFC and amygdala can act in favour both of a passive avoidance and an active approach behavior, like our proposition concerning passive "erasure" versus active "new learning".

It is worth to note that the above three differences between these models are exactly coinciding with the stated in the Introduction three most important mechanisms that are necessary for unified model of extinction and probably of Pavlovian conditioning in general.

Acknowledgment
I would like to thank Hegumen Innokentiy and Olga Jedyaeva for their technical assistance, and John Reynolds, Ludmila Melnikova, and Irina Reynolds for editorial assistance.

References

[1] Barad M, Gean PW, Lutz B, The role of the amygdala in the extinction of conditioned fear, *Biological Psychiatry*, Vol. 60, No. 4, 2006, pp. 322-408

[2] Bevilaqua LR, Bonini JS, Rossato JI, Izquierdo LA, Cammarota M, Izquierdo I, The entorhinal cortex plays a role in extinction, *Neurobiol Learn Mem,* Vol. 85, No. 2, 2006, pp. 192-197

[3] Bouton ME, Context, ambiguity, and unlearning: sources of relapse after behavioral extinction, *Biological Psychiatry*, Vol. 52, No. 10, 2002, pp. 976-986

[4] Bouton ME, Context and behavioral processes in extinction. *Learning and Memory*, Vol. 11, No. 5, 2004, pp. 485-494

[5] Büchel C, Dolan RJ, Armony JL, Friston KJ, Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. *The Journal of Neuroscience*, Vol. 19, No. 24, 1999, pp.10869-10876

[6] Cammarota M, Bevilaqua LR, Vianna MR, Medina JH, Izquierdo I. The extinction of conditioned fear: structural and molecular basis and therapeutic use, *Revista brasileira de psiquiatria*, Vol. 29, No. 1, 2007, pp. 80-5

[7] Corcoran KA, Desmond TJ, Frey KA, Maren S, Hippocampal inactivation disrupts the acquisition and contextual encoding of fear extinction, *The Journal of Neuroscience*, Vol. 25, No. 39, 2005, pp. 8978-8987

[8] Delamater AR, Experimental extinction in Pavlovian conditioning: Behavioural and neuroscience perspectives, *The quarterly journal of experimental psychology*, Vol. 57B No. 2, 2004, pp. 97–132

[9] Gershman SJ, Blei DM, Niv Y, Context, learning, and extinction, *Psychological Review*, Vol. 117, No. 1, 2010, 197–209

[10] Gozzi A, Jain A, Giovanelli A, Bertollini C, Crestan V, Schwarz AJ, Tsetsenis T, Ragozzino D, Gross CT, Bifone A, A neural switch for active and passive fear, *Neuron*, Vol. 67, No. 4, 2010, pp. 656-666

[11] Gray JA and McNaughton N, *The Neuropsychology of Anxiety*, Oxford University Press, Oxford, 2000

[12] Ji J and Maren S, Hippocampal involvement in contextual modulation of fear extinction, *Hippocampus*, Vol. 17, No. 9, 2007, pp. 749-758

[13] Herry C, Ciocchi S, Senn V, Demmou L, Christian Muller , Luthi A, Switching on and off fear by distinct neuronal circuits, *Nature*, Vol. 454, 2008, pp. 600-606

[14] Herry C, Ferraguti F, Singewald N, Letzkus JJ, Ehrlich I, Lüthi A, Neuronal circuits of fear extinction, *European Journal of Neuroscience*, Vol. 31, No 4, 2010, pp. 599–612

[15] Holmes A, Wellman CL. Stress-induced prefrontal reorganization and executive dysfunction in rodents, *Neuroscience and biobehavioral reviews*, Vol. 33, No. 6, 2009, pp. 773-783

[16] Kamprath K, Marsicano G, Tang J, Monory K, Bisogno T, Di Marzo V, Lutz B, Wotjak CT, Cannabinoid CB1 receptor mediates fear extinction via habituation-like processes, *The Journal of Neuroscience*, 2006, Vol. 26, No. 25, pp. 6677-6686

[17] Kamprath K and Wotjak CT, Nonassociative learning processes determine expression and extinction of conditioned fear in mice, *Learning and Memory,* Vol. 11, No. 6, 2004, pp. 770–786

[18] Kim JH, Hamlin AS, Richardson R, Fear extinction across development: the involvement of the medial prefrontal cortex as assessed by temporary inactivation and immunohistochemistry, *The Journal of Neuroscience*, Vol. 29, No. 35, 2009, pp. 10802–10808

[19] Kim JJ, Jung MW, Neural circuits and mechanisms involved in Pavlovian fear conditioning: a critical review, *Neuroscience and Biobehavioral Reviews*, Vol. 30, No 2, 2006, pp. 188–202

[20] Kitazawa S, Neurobiology: Ready to unlearn, *Nature*, Vol. 416, No. 6878, 2002, pp. 270-273

[21] Koenigs M and Grafman J, Post-traumatic stress disorder: The role of medial prefrontal cortex and amygdale, *Neuroscientist*, 2009 Vol. 15, No. 5, pp. 540–548

[22] Kryukov VI, A model of attention and memory based on the principle of the dominant and the comparator function of the hippocampus, *Neuroscience and behavioral physiology*, Vol. 35, No. 3, 2005, pp.35-52

[23] Kryukov VI, The role of the hippocampus in long-term memory: is it memory store or comparator? *Journal of Integrative Neuroscience*, Vol. 7, No. 1, 2008, pp. 117–184

[24] Kumaran D, Maguire EA, Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus*, Vol. 17, 2007, pp.735–748

[25] Kumaran D. Maguire E.A. An unexpected sequence of events: mismatch detection in the human hippocampus. *PLoS biology*. 2006;4:e424

[26] Kumaran D. Maguire E.A. Match mismatch processes underlie human hippocampal responses to associative novelty. *The Journal of Neuroscience,* Vol. 27, No. 32, 2007, pp.8517–8524

[27] Lattal KM, Radulovic J, Lukowiak K, Extinction: does it or doesn’t it? The requirement of altered gene activity and new protein synthesis, *Biological Psychiatry*, Vol. 60, No.4, 2006, pp. 344 –351

[28] Larrauri JA and Schmajuk NA, Attentional, associative, and configural mechanisms in extinction, *Psychological Review*, Vol. 115, No. 3, 2008, pp. 640–676

[29] Lewis PA, Critchley HD, Rotshtein P, Dolan RJ, Neural correlates of processing valence and arousal in affective words, *Cerebral cortex,* Vol. 17, No. 3, 2007, pp. 742-748

[30] Lindsey WC, *Synchronization System in Communication and Control*, Prentice Hall, Englewood Cliffs, NJ, 1972

[31] Mauk MD and Ohyama T, Extinction as new learning versus unlearning: considerations from a computer simulation of the cerebellum, *Learning and Memory*, Vol. 11, No. 5, 2004, pp. 566-571

[32] McNaughton N. The role of the subiculum within the behavioural inhibition system, *Behavioural brain research,* Vol. 174, No. 2, 2006, pp. 232-50

[33] Mcsweeney FK and Swindell S, Common processes may contribute to extinction and habituation, *The Journal of General Psychology*, Vol. 129, No. 4, 2002, pp. 364–400

[34] Muigg P, Hetzenauer A, Hauer G, Hauschild M, Gaburro S, Frank E, Landgraf R, Singewald N, Impaired extinction of learned fear in rats selectively bred for high anxiety--evidence of altered neuronal processing in prefrontal-amygdala pathways, *European neuroscience association,* 2008 Vol. 28, No. 11, pp.2299-2309

[35] Myers KM and Davis M, Behavioral and neural analysis of extinction, *Neuron*, 2002 Vol. 36, No. 4, pp. 567-84

[36] Myers KM and Davis M, Mechanisms of fear extinction, *Molecular Psychiatry*, Vol. 12, No. 2, 2007, pp. 120–150

[37] Myers KM, Ressler KJ, Davis M, Different mechanisms of fear extinction dependent on length of time since fear acquisition, *Learning and Memory*,Vol. 13, No. 2 , 2006, pp. 216–223

[38] Narayanan RT, Seidenbecher T, Kluge C, Bergado J, Stork O and Pape H-C, Dissociated theta phase synchronization in amygdalohippocampal circuits during various stages of fear memory, *European Journal of Neuroscience*, Vol. 25, No. 6, 2007, pp. 1823–1831

[39] Narayanan RT, Seidenbecher T, Sangha S, Stork O and Pape H-C, Theta resynchronization during reconsolidation of remote contextual fear memory, *Behaviour*, Vol. 18 No. 11, 2007, pp. 1107-1111

[40] Pavlov IP, *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford University Press, London, 1927

[41] Pape HC, Narayanan RT, Lesting J, Stork O, Seidenbecher T, Kluge C, Sangha S. Distinctive patterns of theta synchronization in amygdalo-hippocampal-prefrontal cortical circuits during fear memory consolidation and extinction. *Soc Neurosci Abstr 680.8*, 2009

[42] Pape HC, Pare D, Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear, *Psychological Review*, Vol. 90: No. 2, 2010, pp. 419–463

[43] Quirk GJ, Garcia R, González-Lima F. Prefrontal mechanisms in extinction of conditioned fear, *Biological Psychiatry*, Vol. 60, No. 4, 2006, pp. 337-343

[44] Quirk GJ and Mueller D, Neural mechanisms of extinction learning and retrieval, *Neuropsychopharmacology,* Vol. 33, No. 1, 2008, pp. 56–72

[45] Radulovic J and Tronson NC, Molecular specificity of multiple hippocampal processes governing fear extinction, *Review Neuroscience*. 2010; Vol. 21, No.1, pp. 1-17

[46] Redish AD, Jensen S, Johnson A, Kurth-Nelson Z, Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling, *Psychological Review,* 2007, Vol. 114, No. 3, pp. 784-805

[47] Rescorla RA, Wagner AR, A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In: Black A, Prokasy W (eds), *Classical Conditioning*, vol 2: Current Research and Theory. Appleton-Century-Crofts, New York, 1972, pp. 64–99

[48] Robbins SJ, Mechanisms underlying spontaneous recovery in autoshaping, Journal of experimental psychology: *Animal behavior processes*, Vol. 16, No. 3, 1990, 235-249

[49] Sangha S, Narayanan RT, Bergado-Acosta JR, Stork O, Seidenbecher T, Pape HC, Deficiency of the 65 kDa isoform of glutamic acid decarboxylase impairs extinction of cued but not contextual fear memory, *The Journal of Neuroscience,* Vol. 29, No. 50, 2009, pp.15713-15720

[50] Scheltens P, Lopes da Silva FH, Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways. *Hippocampus*, Vol. 10, No. 4, 2000, pp. 398-410

[51] Schiller D and Delgado MR, Overlapping neural systems mediating extinction, reversal and regulation of fear, *Trends in Cognitive Sciences*, Vol. 14, No. 6, 2010, pp. 268–276

[52] Seidenbecher T, Laxmi TR, Stork O, Pape H-C, Amygdalar and hippocampal theta rhythm synchronization during fear memory retrieval, *Science*, Vol. 301, No. 5634 , 2003, pp. 846-850

[53] Siepmann M, and Joraschky P, Modelling Anxiety in Humans for Drug Development, *Current Neuropharmacology*, Vol. 5, 2007, pp. 65-72

[54] Small DM, Gregory MD, Mak YE, Gitelman D, Mesulam MM, Parrish T, Dissociation of neural representation of intensity and affective valuation in human gestation, *Neuron,* Vol. 39, No.4, 2003, pp.701-11

[55] Tronson NC, Schrick C, Guzman YF, Huh KH, Srivastava DP, Penzes P, Guedea AL, Gao C, Radulovic J, Segregated populations of hippocampal principal CA1 neurons mediating conditioning and extinction of contextual fear, *The Journal of Neuroscience*, Vol. 29, No. 11, 2009, pp. 3387-3394

[56] Vianna MR, Igaz LM, Coitinho AS, Medina JH, Izquierdo I, Memory extinction requires gene expression in rat hippocampus, *Neurobiology of learning and memory*, Vol. 79, No.3, 2003, pp. 199-203

[57] Vinogradova OS, Hippocampus as comparator: role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus*, Vol. 11, No. 5, 2001, pp. 578-598

[58] Vinogradova OS, Pribram R.I.I.K.H, The Hippocampus. Vol. 2. Plenum Press; New York: 1975. *Functional organisation of the limbic system in the process of registration of information: facts and hypotheses*; p. 3-69

[59] Watanabe Y, Gould E, McEwen BS, Stress induces atrophy of apical dendrites of hippocampal CA3 pyramidal neurons, *Brain research reviews*, Vol. 588, No. 2, 1992, pp.341-345

[60] Witter MP, Naber PA, van Haeften T, Machielsen WCM, Rombouts S, Barkhof F, Scheltens P, Lopes da Silva FH, Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways, *Hippocampus*, Vol. 10, No. 4, 2000, pp. 398-410

[61] Yamaguchi S, Hale LA, D'Esposito M, Knight RT, Rapid prefrontal-hippocampal habituation to novel events, *The Journal of Neuroscience,* Vol. 24, No. 7, 2004, pp. 5356–5363