Intrinsic volatility of synaptic connections — a challenge to the synaptic trace theory of memory
Gianluigi Mongillo\textsuperscript{1,2}, Simon Rumpel\textsuperscript{3} and Yonatan Loewenstein\textsuperscript{4}

According to the synaptic trace theory of memory, activity-induced changes in the pattern of synaptic connections underlie the storage of information for long periods. In this framework, the stability of memory critically depends on the stability of the underlying synaptic connections. Surprisingly, however, synaptic connections in the living brain are highly volatile, which poses a fundamental challenge to the synaptic trace theory. Here we review recent experimental evidence that link the initial formation of a memory with changes in the pattern of connectivity, but also evidence that synaptic connections are considerably volatile even in the absence of learning. Then we consider different theoretical models that have been put forward to explain how memory can be maintained with such volatile building blocks.

### Synapses and memory

A needle that punctures a cloth leaves a small mark, making it easier to re-puncture that point. René Descartes, the 17th century philosopher, mathematician and scientist used this metaphor (see Figure 1a) to explain his theory of memory that was based on animal spirits flowing through pores in the brain [1]. This theory is long obsolete. However, the metaphor still nicely captures the very essence of the contemporary synaptic trace theory of memory. Today, it is generally believed that specific patterns of synaptic connections in the brain are a major physical correlate of long-term memories. Changes in the synaptic strength between a pair of neurons are driven by their specific patterns of electrical activities and neuromodulators, which are in turn elicited by behaviorally relevant experiences [2] (Figure 1b). The needle in this framework is the neural activity, whereas the specific pattern of synaptic connections are a manifestation of the marks in the cloth [3].

The maturation of transgenic techniques, which enable us to label single neurons in the living brain and two-photon microscopy, which allows us to image these neurons with a submicron resolution have provided support to the synaptic trace theory of memory. Specifically, we are now able to chronically image cortical spines, small protrusion on the dendritic shaft, on which most of neocortical excitatory synapses reside. In acute slices, spines sense synaptic activity within 30 min of their formation [4], and combined in vivo imaging and post hoc serial sectioning electron microscopy has shown that spine structures that are older than four days show the characteristic morphological features of a functional synapse [5]. Therefore, the formation and elimination of a spine are good proxies to the formation and elimination of a synaptic connection, and spine volume can serve as proxy for the functional strength of a synaptic connection [6–9]. Indeed, it has long been known that tetanic stimulation results in an increase in the number of spines [10,11] and an enlargement of their size [12]. More recently, learning has been demonstrated to be correlated with a transient increase in the density of spines [13,14,15,16]. Finally, a causal link between memory formation and synaptic changes has been established. The specific erasure of the spines formed during training results in a specific erasure of the corresponding memory [17]. In a different study, optogenetics was used to encode a specific fear memory and then to induce long-term depression and long-term potentiation within the circuit to remove and to reinstated this memory [18]. Whereas most studies focused on the excitatory connections, there is also evidence supporting a role of inhibitory plasticity in memory formation [19,20].

### Synaptic volatility challenges the synaptic trace theory

If a specific pattern of connectivity is the physical basis of long term memory, then our ability to retrieve this
memory after years, and even decades, crucially entails that this pattern remains stable. Surprisingly however, cortical spines are being constantly formed and eliminated in the living brain, implying that the corresponding synapses are highly volatile (Figure 2a,b). For example, a recent study suggested that within several weeks, the synaptic network in the hippocampus CA1 region is completely remodeled [21]. In the auditory cortex, 60% of the spines are replaced within three weeks [22] (Figure 2c). Somewhat lower turnover rates have been reported for different cortical regions [23–25]. The substantial cortical volatility manifests not only in the high rates of turnover but also in substantial changes in the sizes of the spines [25,26], which is indicative of changes in the efficacies of the corresponding excitatory synapses that resides on them [27–29]. Specifically in the auditory cortex, 70% of the spines that are not eliminated change their size by at least a factor of two within three weeks [30] (Figure 2d).

Another challenge to synaptic trace theory is the finding that synapses change in the absence of neural activity: according to the synaptic trace theory of memory, changes in connectivity are driven by specific, experience-dependent neural activities. In this framework, blocking the activity should abolish these changes. However, in vitro studies have challenged this hypothesis. The volumes of spines in neuronal cultures fluctuate substantially. Surprisingly, blocking neuronal activity, as well as NMDA-sensitive glutamate receptors does not prevent these fluctuations but only decreases their frequency and magnitude [31]. Similarly, the sizes of the post-synaptic densities (PSDs) substantially fluctuate over time even when network activity is completely blocked [32].

The relevance of these results to the living brain has been recently demonstrated in a study, in which calcium inhibitors, as well as sodium channel blockers were infused into the visual cortex of adult mice. While these manipulations decreased the rate of spine turnover in mice reared in visually enriched environmental conditions, they had almost no effect on the rate of spine turnover in normal conditions [33]. Together, these observations suggest that synaptic connections follow intrinsic

---

Figure 1

Trace theories of memory. (a) This picture was used by Descartes to illustrate his theory of memory: ‘... if one were to pass several needles or engraver’s points through a linen cloth [. . .] they would leave traces in the cloth, which would stay open [. . .] or [. . .] make them very easy to open again’ (see p. 69 of [1]). The needles are the experience-induced activity that produce the stable structural modifications, the corresponding marks in the cloth. (b) The synaptic trace theory of memory. Initially, all synaptic connections are weak (thin lines) (†). An experience selectively activates a subpopulation of neurons (red), which through activity-dependent synaptic plasticity strengthen the corresponding synaptic connections (red lines) (2). These strengthened connections (thick lines) remain stable (3) and allow the retrieval of the memory by supporting a pattern of neural activity similar to the one elicited by the original experience (4).
Volatility of dendritic spines of pyramidal neurons in the mouse auditory cortex. **(a)** Six images of the same dendrite, imaged at an interval of four days. Solid circles denote spines that were present in all imaging sessions, whereas open circles denote spines, found only in some imaging sessions. **(b)** The sizes of all spines in **(a)**, color-matched. **(c)** Lifetime plot of 3688 imaged spines. Presence of a spine is indicated by a thin horizontal black line on a given day. For example, spines 1–1420 are those observed in the first imaging session; spines 1421–1937 are those observed in at least the first and the second imaging session; spines 1938–2433 are those observed for the first time in the second imaging session; and so on. **(d)** The sizes of 100 randomly-chosen spines present throughout the experiment in the six imaging sessions, color-coded. Data was reproduced from [22,30] and is available online at http://bio.huji.ac.il/yonatanlab/spines/.

Dynamics that lead to their continuous remodeling in the living brain on top of which activity-dependent changes may occur.

It is well established that on average, pyramidal neurons in the cortex are connected by several synapses [34–37]. Therefore, another method of estimating the contribution of activity-independent processes to synaptic changes is to consider synapses that 'share' the same presynaptic and postsynaptic neurons. If the activities of the presynaptic and postsynaptic neurons solely and deterministically govern the changes in the efficacies of the corresponding synapse, then such synapses should be identical. Indeed, when two synaptic contacts of the same connection between a pair of pyramidal neurons in the somatosensory cortex slices were examined, their release probabilities were similar [35]. Furthermore, three dimensional reconstructions from serial electron microscopy of pairs of spines of the same neuron that are contacted by the same axon in the somatosensory cortex of adult mice [36] and the hippocampus of adult rats [38] tended to be more similar than randomly-chosen pairs [36,37]. Intriguingly, however, as noted in [39], the synapses were far from identical and spine volumes in these pairs differed by a factor of two. Similarly, a recent study considered PSD sizes of commonly innervated and non-commonly innervated synapses in neuronal cultures [39]. The authors concluded that less than 40% of glutamatergic synapse size remodeling could have been attributed to specific activity histories. Thus, the contribution of activity-independent processes to the efficacy of the synapse is at least comparable to the activity-dependent component.

Finally, a recent in vivo study demonstrated that the longevity of a spine can be predicted, to some extent from its morphological features [22]. The idea behind
this study was that the more a change in a synapse can be predicted, the less likely it is that it reflects the acquisition of a new memory pattern or a skill, unknown at the time of prediction.

In summary, while recent years bring direct support to the synaptic trace theory, other experimental results challenge it in two ways. First they demonstrate that a substantial fraction of the changes in synaptic efficacies are not driven by the electrical activities of the corresponding neurons. In the framework of the synaptic trace theory of memory, these activity-independent changes should be considered as ‘noise’. Second, they point out a potential inconsistency between the timescales of synapses and memories, namely, that information stored in synapses is degraded in timescales much shorter than the degradation of long-term memories. Synaptic noise can under specific circumstances be viewed as a feature rather than a bug when considering the fact that direct gradient policy learning [40,41], a standard reinforcement learning algorithm, can be achieved by randomly perturbing synaptic connections [42,43]. However, it is clear that memory cannot be maintained if all its physical traces are lost [44].

Memory in the presence of synaptic volatility

In the remainder of this review we will consider three types of solutions that have been proposed to explain how functional stability is maintained in a volatile brain. We will focus on synaptic mechanisms, bearing in mind the possibility that stable, non-synaptic elements may also contribute to the storage of information for long periods [45,46].

One explanation to the apparent discrepancy between the timescales of synaptic turnover and the timescales of memory maintenance is — that there is no discrepancy. The timescale of some biophysical elements is, in fact, as long as the longest memory timescale (Figure 3a). For example, it has been shown that a small fraction of spines is stably maintained in mice over many months [167] suggesting that long-term memories are maintained in these specifically stable spines. A more quantitative analysis of spine turnover has shown that there are no two classes of spines, stable and transient. Rather, spines continuously stabilize with their age at multiple time scales and a small fraction of the spines is predicted to remain stable as long as the longest memories [227]. In line with these findings of multiple time scales, theoretical studies have demonstrated that multiple timescales in the dynamics of synapses substantially enhance their ability to store a large number of memory patterns [47–49].

Most of what is currently known about the dynamics of dendritic spines in a limited number of types of synaptic connections as proxies for excitatory synapses. Another possibility worthwhile considering is that plasticity of excitatory synapses is essential for the formation of long-term memories, but that the long term maintenance is mediated by other, more stable synapses. A recent theoretical study has demonstrated that despite their smaller number, the pattern of activity of cortical networks is dominated by inhibitory and not by excitatory synapses, although they comprise a much larger fraction of all synaptic connections. Therefore, the capacity of the inhibitory network to store memories is surprisingly larger than that of the excitatory network [50]. Technical developments in recent years enabled now chronic monitoring of inhibitory synapses in vivo and have provided evidence for plasticity in the inhibitory network as well [51–55].
The specific link of this type of plasticity to various forms learning and memory is still to be investigated.

A second explanation to the discrepancy between spine volatility and memory stability is that memory is maintained in connections between two neurons and not in individual synapses (Figure 3b). Pyramidal neurons in the cortex that are typically connected by several synapses [34–37]. Therefore, connections between cells can be stable and underlie long-term memories even in the presence of substantial synaptic volatility, provided that synaptic changes belonging to the same pair of neurons co-vary such that the overall connection strength remains unchanged. Such a solution, however, requires correlation of synaptic changes at the level of a single neuron [56]. Several recent models have proposed a mechanism that is based on activity-dependent plasticity for maintaining such connection-stability in face of synaptic volatility [57–59]. For example, according to one model [58], each potential contact is randomly transformed into a synapse and the efficacies of these synapses and their elimination are driven by spike-timing dependent plasticity. If a pair of neurons is connected by several synapses, then the correlation in the activities of the presynaptic and postsynaptic neurons will facilitate those synapses and thus stabilize the newly-formed synapses. This correlation-driven positive-feedback can maintain the existence of connections for long periods of time even when the underlying synapses are unstable. Note however, that in these dynamics that combine positive feedback and saturation, analog information about the strength of the connection cannot be maintained [60]. Due to technical limitations, this hypothesis has not been put to an experimental test yet.

Finally, memories can be stabilized at the network level [61]. If memory patterns are stored in the synapses of an attractor neural network [62], as in the Hopfield model [63], then perpetually revisiting the attractor states, combined with Hebbian synaptic plasticity can stabilize the memory patterns [64,65] (Figure 3c). However, this framework requires that all memories will be repeatedly (and subconsciously) visited within the timescale of synaptic-trace decay [66] (but see [67]). An interesting theoretical study has demonstrated that spike-timing dependent plasticity can also stabilize memories that are not explicitly visited. The idea behind this intriguing study is that in the presence of noise, the correlation structure of the neuronal activity bears a signature of the entire connectivity matrix and thus can enhance it [68*]. However, more work is needed in order to determine whether this mechanism is sufficient to stably maintain a large number of memory patterns by unstable synapses.

Concluding remarks
Here we reviewed some of the recent evidence supporting the synaptic trace theory of memory. This theory agrees with our everyday intuition that information is stored for long-periods of time in elements that are physically as stable as possible, be it an inscription on an obelisk or the needle marks in Descartes’ cloth. Yet, we pointed out a substantial challenge to this theory, the fact that the ‘stable’ elements of this theory — the synapses, are in fact highly volatile. We discussed some attempts to reconcile this volatility with stability. However, it is still unclear at what level stability is achieved in the brain. Thus, the question of how the brain maintains functional stability with volatile elements remains a fundamental puzzle in neuroscience.

Conflict of interest statement
Nothing declared.

Acknowledgements
This work was supported by the Israel Science Foundation (Grant No. 757/16), the DFG (CRC 1080) and by the Gatsby Charitable Foundation. We thank David Hansel, Noam Ziv and Haruo Kasai for carefully reading the manuscript and helpful comments and Zehava Cohen for help in preparing Figure 1.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:
• of special interest
•• of outstanding interest


   One of the first studies (see also Yang et al., 2009) demonstrating a direct link between behavioral learning and spine formation. Successful learning is shown to induce rapid increase in spines’ density followed by spine pruning, preferentially preserving learning-induced spines, that eventually restores density.
   The first study demonstrating a causal link between behavioral learning and spine formation. Using a photosensitive protein, selectively expressed by newly formed spines, the authors show that the elimination of spines that were induced during learning disrupts behavioral performance.
   The authors demonstrate that, in absence of learning, the longevity of spines can be quantitatively predicted using their age and their morphological features. The study paves the way to a precise quantification of learning-induced changes in spines’ dynamics.

In vivo study demonstrating that the majority of baseline spines’ turnover is independent of neuronal activity. This result strongly supports the idea that a substantial fraction of the spontaneous spines dynamics observed in vivo do not reflect latent learning processes.
   The authors compare the covariation of PSD sizes of synapses that share the same pre-synaptic and post-synaptic neurons to those formed by different presynaptic neurons in cultured cortical networks. Common pre-synaptic and post-synaptic activity histories account for approximately half of the covariation, suggesting that activity-independent processes substantially contribute to changes in synaptic efficacies.
50. Mongillo G, Rumpel S, Loewenstein Y: Synaptic volatility and the reorganization of electrical activity in neuronal networks. COSYNE. 2015, Salt Lake City, UT.
68. Wei Y, Koulakov AA: Long-term memory stabilized by noise-induced rehearsal. J Neurosci 2014, 34:15804-15815. Memories can be maintained if they are frequently reactivated, so that the corresponding synaptic structure can be repaired. This intriguing theoretical study suggests that neuronal noise, in combination with suitable synaptic plasticity mechanisms, could prevent the decay of a stored memory, even if this memory is not reactivated.