Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Does mental context drift or shift?

Sarah DuBrow[1], Nina Rouhani[1,2], Yael Niv[1,2] and Kenneth A Norman[1,2]

Theories of episodic memory have proposed that individual memory traces are linked together by a representation of context that drifts slowly over time. Recent data challenge the notion that contextual drift is always slow and passive. In particular, changes in one's external environment or internal model induce discontinuities in memory that are reflected in sudden changes in neural activity, suggesting that context can shift abruptly. Furthermore, context change effects are sensitive to top-down goals, suggesting that contextual drift may be an active process. These findings call for revising models of the role of context in memory, in order to account for abrupt contextual shifts and the controllable nature of context change.

Addresses
[1] Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, United States
[2] Department of Psychology, Princeton University, Princeton, NJ 08544, United States

Corresponding author: Norman, Kenneth A (knorman@princeton.edu)

Current Opinion in Behavioral Sciences 2017, **17**:xx–yy

This review comes from a themed issue on **Memory in time and space**

Edited by **Lila Davachi** and **Neil Burgess**

http://dx.doi.org/10.1016/j.cobeha.2017.08.003

Introduction

Much of what we experience is transient. Yet, many of our internal representations tend to show inertia, drifting more gradually than our external environment. Slowly drifting context can act as a bridge between more transient item representations. That is, if two items are associated in memory to the same slowly changing context, this provides an indirect route for one of those items to activate the other [1]. This property has been captured in computational models of temporal context [2–5], which posit that *gradual integration of sensory inputs* creates a slowly drifting context representation that can then be linked to representations of individual items. These models have been very successful in explaining psychological and neural data, but they have recently been challenged by data suggesting that mental context can rapidly shift in response to both surprising events and changing task demands. Below, we first review data consistent with the context drift framework. Then, we describe three lines of challenging findings, as well as theories that have the potential to reconcile the 'drifting' and 'shifting' nature of mental context.

Evidence for slow, passive contextual drift

When freely recalling random lists of words, two effects have been identified as particularly strong and reliable — the recency effect and the contiguity effect — both of which can be explained by a slowly drifting mental context representation bound to items in memory [2]. The recency effect, or the tendency to recall end-of-list items especially well, can be explained by the idea that context in the recall period is most similar to context at the end (vs. beginning) of the list. The contiguity effect, or the tendency to transition in recall between items that were close to each other at study, can be explained by the idea that neighboring items from the study period share context, and therefore prime each other for recall. Neural data, particularly in the hippocampus, have also shown properties consistent with slow and automatic drift. Hippocampal 'time cells' show gradually changing activity patterns on the order of seconds during unfilled delays even when the animal's location is fixed [6–9], consistent with automatic drift. Sensitivity to longer timescales on the order of days to weeks has also been observed in hippocampal place cells [10–12] and independently of place coding [13,14]. This neural drift may influence memory by enabling distinct events encountered in close temporal proximity to be linked according to the overlap in their neural activity profiles [15]. Furthermore, the degree to which neural activity patterns change across events has been associated with memory for temporal order and distance [16–19]. Critically, this relationship has been observed in the same regions that show slow drift [20••,21]. A link between slow drift and spontaneous organization of memory was recently established in an fMRI study that tracked lingering activation of recently experienced items [22]. When neural activity associated with the previous item's category (celebrity, location, or object) persisted into the current item's encoding period, this led to items being clustered at recall according to the previous item's category (as would be expected if this slowly drifting category activity served to contextualize the current item's memory trace).

Challenging data: can context shift abruptly?

Recent work has challenged the notion that contextual drift is always gradual. In particular, inducing abrupt changes in stimulus features and/or task goals creates separation in memory [19,23–25]. Similar effects can also be observed with naturalistic stimuli such as written stories or films, in which changes ('event boundaries') occur at the narrative level (e.g., going from cooking a meal to eating the meal) [26]. Event boundaries can exert a sharp disruptive effect on memory, such that accessing information across a boundary is impaired even when controlling for the time elapsed [27–29]. Cognitively, these findings have been explained in terms of the idea that participants form *situation models* that describe the properties of the event [30,31]. One's currently active situation model may be a particularly strong component of mental context, and thus changes in that model may serve as powerful context shifts.

Understanding how situation models might be implemented in the brain has been a major focus of recent work. Researchers have argued that a posterior medial network (PMN), including the parahippocampal cortex, retrosplenial cortex, and other regions of the default network, integrates internal and external information in order to represent the features of the current event [32]. This proposal is consistent with data showing that these regions are capable of integrating information across long time scales on the order of minutes [33] — an important prerequisite for constructing models of events that unfold at this time scale. This same network has been found to represent information about specific movie scenes during both movie viewing and memory recall, even across individuals [34]. A key property of brain regions involved in representing situation models is that neural patterns should be relatively stable within a temporally extended event and change abruptly at event boundaries. By fitting a model with this property to movie-viewing fMRI data, Baldassano and colleagues [35••] showed that regions in the PMN are well-described by shifts between relatively stable patterns, lasting on the order of minutes. The 'shift' moments identified in the PMN corresponded to moments that human observers identified as event boundaries, further suggesting that these regions are sensitive to high-level event content. Interestingly, perceptual regions tended to identify a larger number of short events, presumably due to a greater sensitivity to rapidly changing environmental features. In a different study, the hippocampus was found to be sensitive to narrative context at an even coarser level than the PMN, separately representing two interleaved story lines [36•]. Thus, the hippocampus and PMN may support high-level structure, integrating within stable events and rapidly shifting when a new situation is encountered. Together with lower-level regions, multiple timescales or hierarchies of event structure could be represented simultaneously.

Inferring change in the state of the world

How can computational models account for these sharp shifts in mental context? Within the temporal context framework, an abrupt increase in contextual drift rate can explain boundary effects on memory [37], but what might drive this increase? One possibility is that prediction errors, or discrepancies between expected and actual outcomes, generate event boundaries and ensuing context shifts [38–40], consistent with evidence that prediction errors can drive segmentation in learning [41,42]. According to a recent theory of latent cause inference, people build up a library of statistical models of their environment, and then infer which is the hidden, or 'latent,' cause of their current sensory observations [43,44•,45,46]. Large prediction errors signal that the current latent cause is no longer relevant (i.e., the situation has changed) and a new or stored latent cause should be inferred. Figure 1 illustrates how this inference process might shape mental context during a real-world situation: a surprise party at work. The appearance of a birthday cake during a meeting elicits a prediction error signaling that work is no longer relevant; instead, the new latent cause is a party.

One way prediction errors may influence memory is by segmenting experience when the inferred latent cause is updated, thereby reducing interference. In support of this idea, Gershman and colleagues [47] found that old items (simple line segments) were better remembered when they were followed by a prediction error and therefore 'separated' from similar, interfering, items. Rouhani and colleagues [48] further showed that large reward prediction errors boosted performance on multiple memory measures, consistent with the idea that prediction errors led to increased segmentation and reduced interference. By contrast, when changes are small and predictable, a single latent cause is inferred and memories are more malleable and susceptible to interference [49,50]. Neurally, the orbitofrontal cortex, which is sensitive to unobservable states [51], has been shown to represent the probability that each latent cause is currently active during the inference process [52]. Thus, one intriguing possibility is that the orbitofrontal cortex may draw on memory representations from the hippocampus and temporally extended situational information from the PMN to represent the current context for goals and actions.

Goal-directed control of context

While latent cause models can account for abrupt context shifts, they do not (in their most basic form) speak to the strong influence of people's goals on the extent to which context lingers. For example, in much of the event segmentation literature, participants are asked to indicate when an event boundary occurs. This focus on detecting changes is associated with substantial improvements in memory performance [53,54,55•]. One potential explanation for this improvement is that the task of segmenting