Short-term memory: $\tau \approx 10$ secs

Classic example: looking up a phone #, remembering it for a few secs until you dial it.

Crucial to everyday life (e.g., conversation)

Onset is very fast -- too fast to be subserved by changes in gene expression or cell structure
Instead, it is thought that it is observed rapid changes in *neural activity* that maintain STM:

A *transient stimulus* leads to *persistent, stimulus-dependent* neural activity.

It’s memory because we can read stimulus id from firing rates, *after stimulus is gone*. Circuit can settle into either stable activity mode A or mode B.

Thus the same neural circuit (same neurons, same connections) must be capable of supporting several different, *stable* activity patterns (τ~seconds)
What is this “stability” you speak of? Stable in the face of what? -- noise, decay

A geometric view (visualization aid) from dynamical systems theory.

Imagine that the dynamics can be described such the state of the system always goes downhill in some imaginary function $L$

$$L = L(r_1, r_2, Ca_1^{++}, \ldots)$$

If we are inside the lip of a well in $L$, we go down and stay forever in the well: we are stable to noise

(near a stable pt, can always find such a description)

Stable pts are “attractors”
We’ve talked about memory for one out of a discrete set of items.

But sometimes you can remember one out of a continuous range of possibilities.
Example neuron from PFC of macaque

Wang, ... Arnsten, Cell 2007

Patricia Goldman-Rakic, 1980s
Old picture can only do one of two discrete memories

Can make a "ring" attractor -- a type of continuous attractor.

Notice translational symmetry along bottom of circular valley
How do you build a ring attractor? easy: short-term excitation, longer-range inhibition

Machens and Brody, *Neural Comp.*, 2008

Where’s the symmetry here?

Ring models also used for head direction (K. Zhang); and perception, e.g. orientation tuning in V1 (Sompolinsky)
Continuous attractors are used as models in many contexts—e.g., grid cells.

Continuous memories don’t have to be in a ring:

Graded memory of $f_1$

We can make a “line” attractor:

Again, notice *translational symmetry* along bottom of valley
Ok, how do we build a line attractor?

Oldest idea about STM comes from Lorente de No in the 1920s: “reverberating activity” in *networks of neurons* can lead to persistent activity after a transient stimulus.
input at time t

input at time t+1

w

output at time t

45° degree line

input->output fn

input at time t

at low w
input at time t

input at time t+1

input->output $f^n$

at high $w$

45° degree line

input at time t
The diagram shows a graph with the input at time $t$ on the x-axis and the input at time $t+1$ on the y-axis. The graph includes a $45^\circ$ degree line, indicating that the function $f^n$ is applied almost right at $w$. The function's output at time $t+1$ is directly proportional to its input at time $t$. The diagram illustrates the relationship between the input and output, emphasizing the linear nature of the function.
\[ x = \text{firing rate} \]

\[ \tau \dot{x} = -x + f(x) \]

\[ \tau \dot{x} = -x + wx \]

\[ \dot{x} = \frac{w - 1}{\tau} x \]

Must fine-tune \( w \)!

If \( \tau \sim 100 \text{ ms} \), then for 10 sec robustness, need \( w \) within 1% of 1

The trouble with any truly continuous attractor: sensitivity to noise in

(1) dynamics

(2) construction

Get only a few stable points, not a continuum
The Koulakov and Lisman approach  (Nat. Neurosci., 2002)

How do computers do it?

If the many \textit{discrete} stable pts are closely spaced enough, the possible stable levels of firing rate appear continuously graded!

Both robustness in dynamics and robustness in construction achieved
The Koulakov ... Lisman discretizing approach (*Nat. Neurosci.*, 2002)

As presented by them using networks:
Previous arguments about continuous attractors are geometric—
they don’t depend mechanism

They apply to ALL mechanisms
(single cell, biochemical, network, etc.);
and ALL timescales, too.

All-or-none potentiation at CA3-CA1 synapses

CARL C. H. PETERSEN*, ROBERT C. MALENKA†‡, ROGER A. NICOLL*§, AND JOHN J. HOPFIELD¶

Departments of *Cellular and Molecular Pharmacology, †Physiology, and §Psychiatry, The University of California, San Francisco, San Francisco, CA, 94143; and ¶Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by Roger A. Nicoll, January 28, 1998

Building graded (analog) memories is hard: discretized memories are easier
Four more views

* Cellular basis (at single-cell level) for persistent activity
  

* Could it all be in short-term dynamics of *synapses* (not firing rate)?
  

* Short-term memory in functionally feedforward networks.
  

* Short-term memory in sequences of neural states.
  
Synaptic transmission blocked

Egorov ... Alonso, Nature 2002
Really a continuum? Sure looks like it
Very long persistence time

AND it’s robust to quick perturbations
Cell with many Ca-dependent channels,
some open | some closed

Firing rate ~ # open channels

Fransen, ... Alonso, Neuron 2006

channels open/close depending on [Ca]

Same idea! -- discretize in # of channels
* Cellular basis (at single-cell level) for persistent activity
  

* Could it all be in short-term dynamics of *synapses* (not firing rate)?
  

* Short-term memory in functionally feedforward networks.
  

* Short-term memory in sequences of neural states.
  
hard decisions are slow, easy decisions are fast

- social decisions ("who do I talk to?") (e.g., Krajibich 2012)
- sensory decisions ("are the dots going left or right?") (e.g., Newsome, 1989)
- economic decisions ("which investment do I make?") (e.g., Gluth 2012)
- gambling decisions ("which number will I bet on?") (e.g., Busemeyer, 1993)
- memory decisions ("what does this remind me of?") (e.g., Ratcliff, 1978)
- visual search decisions ("where is the object I’m looking for?") (e.g., Purcell, 2010)
- value decisions ("which ice cream do I get?") (e.g., Milosavljevic 2012)

common core process, **noisy accumulation of evidence**, can account for all of them

decision-making’s most common behavioral observation

Using sensory stimulus, try to decide between reluctant LEFT or resolute RIGHT.

- weak evidence (hard, slow decision) $a < 0 \Rightarrow$ decide LEFT
- strong evidence (easy, fast decision) $a > 0 \Rightarrow$ decide RIGHT

features that make you lean RIGHT add to $a$,
features that make you lean LEFT subtract from $a$
hard decisions are slow, easy decisions are fast

- social decisions ("who do I talk to?")
- sensory decisions ("are the dots going left or right?")
- economic decisions ("which investment do I make?")
- gambling decisions ("which number will I bet on?")
- memory decisions ("what does this remind me of?")
- visual search decisions ("where is the object I’m looking for?")
- value decisions ("which ice cream do I get?")

applies to
common core process, **noisy accumulation of evidence**, can account for all of them

\[ a > 0 \implies \text{decide RIGHT} \]
\[ a < 0 \implies \text{decide LEFT} \]
Neurons with persistent stimulus-dependent activity usually also have *time*-dependent activity.

for the network to hold the memory of some stimulus $m$, we just need firing rates $r_i$ to be such that there exists some function $g()$ such that

$$g(r_i(m,t)) = m$$

for $0 < t < \sim$ a few secs.

In each behavioural context, the curves are fits of a behavioural model. Performance is shown as a function of motion (left) or colour (right) coherence colour contexts (bottom), averaged over 80 recording sessions (163,187 trials).

Selection and integration underlying behaviour in our task. Responses provide fundamental constraints on the mechanisms of represented in PFC. In particular, four properties of the population (Extended Data Fig. 3) and at the level of individual principal component (motion or colour). We estimated this task-related subspace in two evidence, the strength and direction of the colour evidence, and context subspace that captures across-trial variance due to the choice of the monkeys' behaviour, we represent population features (that is, patterns of activations) in the population response. To projection of the population we used linear regression to define the four orthogonal, task-related axes (Fig. 1).

Evidence, the strength and direction of the colour evidence, and context...
Extended Data Figure 3

Mixed representation of task variables in PFC.

a, Example responses from six well-isolated single units in monkey A. Each column shows average normalized responses on correct trials for one of the single units. Responses are aligned to the onset of the random-dot stimulus, averaged with a 50-ms sliding window, and sorted by one or more task-related variables (choice, motion coherence, colour coherence, context). The green lines mark time intervals with significant effects of choice (a), motion coherence (b), colour coherence (c), or context (d) as assessed by multi-variable, linear regression (regression coefficient different from zero, \( P < 0.05 \)).

Linear regression and coefficient significance are computed over all trials (correct and incorrect, motion and colour context; Supplementary Information, section 6.3). The horizontal grey line corresponds to a normalized response equal to zero.

a, Responses sorted by choice (solid, choice 1; dashed, choice 2) averaged over both contexts.
b, Responses during motion context, sorted by choice and motion coherence (black to light-grey, high to low motion coherence).
c, Responses during colour context, sorted by choice and colour coherence (blue to cyan, high to low colour coherence).
d, Responses sorted by choice and context (black, motion context; blue, colour context). As is typical for PFC, the activity of the example units depends on many task variables, indicating that they represent mixtures of the underlying task variables.

e, f, De-noised regression coefficients for all units in monkey A (e) and monkey F (f).

The data in Extended Data Fig. 1 are re-plotted here to directly compare the effects of different task variables (choice, motion, colour, context) to each other. Each data point corresponds to a unit, and the position along the horizontal and vertical axes is the de-noised regression coefficient for the corresponding task variable. The horizontal and vertical lines in each panel intersect at the origin (0,0). Scale bars span the same range (0.1) in each panel. The different task variables are mixed at the level of individual units. Although units modulated by only one of the task variables do occur in the population, they do not form distinct clusters but rather are part of a continuum that typically includes all possible combinations of selectivities. Significant correlations between coefficients are shown in red (\( P < 0.05 \), Pearson's correlation coefficient \( r \)).

Extended Data Figure 3 | Mixed representation of task variables in PFC. 

(a–d), Example responses from six well-isolated single units in monkey A. Each column shows average normalized responses on correct trials for one of the single units. Responses are aligned to the onset of the random-dot stimulus, averaged with a 50-ms sliding window, and sorted by one or more task-related variables (choice, motion coherence, colour coherence, context). The green lines mark time intervals with significant effects of choice (a), motion coherence (b), colour coherence (c), or context (d) as assessed by multi-variable, linear regression (regression coefficient different from zero, \( P<0.05 \)). 

Linear regression and coefficient significance are computed over all trials (correct and incorrect, motion and colour context; Supplementary Information, section 6.3). The horizontal grey line corresponds to a normalized response equal to zero.

(a), Responses sorted by choice (solid, choice 1; dashed, choice 2) averaged over both contexts.

(b), Responses during motion context, sorted by choice and motion coherence (black to light-grey, high to low motion coherence).

(c), Responses during colour context, sorted by choice and colour coherence (blue to cyan, high to low colour coherence).

(d), Responses sorted by choice and context (black, motion context; blue, colour context). As is typical for PFC, the activity of the example units depends on many task variables, indicating that they represent mixtures of the underlying task variables.

(e), (f), D e -noised regression coefficients for all units in monkey A (e) and monkey F (f). The data in Extended Data Fig. 1 are re-plotted here to directly compare the effects of different task variables (choice, motion, colour, context) to each other. Each data point corresponds to a unit, and the position along the horizontal and vertical axes is the de-noised regression coefficient for the corresponding task variable. The horizontal and vertical lines in each panel intersect at the origin (0,0). Scale bars span the same range (0.1) in each panel. The different task variables are mixed at the level of individual units. Although units modulated by only one of the task variables do occur in the population, they do not form distinct clusters but rather are part of a continuum that typically includes all possible combinations of selectivities. Significant correlations between coefficients are shown in red (\( P<0.05 \), Pearson's correlation coefficient \( r \)).
The first model (Fig. 3b) is based on two widely accepted hypotheses: the evidence selection model and the evidence combination model. First, it assumes that inputs are selected early in the decision process, as suggested by our PFC recordings (Fig. 3a). From this, it follows that inputs are evaluated individually from each other (Fig. 3b–d), and can thus be validated or rejected by mechanisms such as the combination of sensory inputs. Figure 2 depicts this process, where the motion input directly elicits a pattern of activation in PFC resembling the expected pattern corresponding to a choice (the grey arrow in Fig. 3b, top), but is filtered out before reaching PFC when irrelevant inputs are present (no grey arrow in Fig. 3b, bottom). Second, it assumes that the relevant direction of the momentary evidence and the axis of choice determines how much the corresponding input affects behaviour. Furthermore, the actual momentary behaviour (Fig. 1e), whereas the latter has a large behavioural effect (Fig. 2a, dark grey). Yet the former has almost no behavioural effect (Fig. 2d, black) elicits a larger deflection along the motion axis (relative to the choice axis), but is filtered out before reaching PFC when irrelevant inputs are present (Fig. 3d) lend little support for variable choice (Fig. 3d) or input (Fig. 3c) axes. More generally, the PFC data from monkey A rule out any model of integration for which the degree of overlap between the relevant and irrelevant inputs is critical. This is based on the observed PFC responses along the choice and motion axes (Extended Data Fig. 3–8), such that the population response for a given condition and time is represented as a point in a field of the recorded neurons. The direction of the colour input does not refer to the motion input (blue colours). Here, choice 1 corresponds to the target in the response field of identity (motion context, top; colour context, bottom). At the single neuron level, variable axes that change across contexts would be reflected as complex, nonlinear interactions between context and the other task variables, which have been proposed in some task-switching models. In the task-switching models, the angle of the motion input in the motion context, projected onto the axes of choice and motion, reveals the projection onto the axis of colour. The observed PFC responses also rule out two additional models of evidence. The average sensory stimulus points towards the chosen target (3 conditions per chosen target); for irrelevant inputs (Fig. 3c), however, the stimulus can point either towards or away from the chosen target on correct trials (6 conditions per key feature observed in monkey A. Most importantly, the observed PFC responses also rule out two additional models of evidence integration.
Trained network; “unfolding” in time