

Supporting Information

Hopfield 10.1073/pnas.0913991107

SI Modeling and Mathematics

The equations of motion used for the adapting integrate-and-fire neurons during the time between action potentials of neuron k are

$$du_k/dt = -u_k/\tau_{\text{cell}} + I_{\text{bias}} + I_{\text{synaptic}} + I_{\text{sensory}} - I_{\text{global}} - J_k \quad \text{[S1]}$$

$$dJ_k/dt = -J_k/\tau_{\text{unadapt}}. \quad \text{[S2]}$$

J_k represents an inward inhibitory current that could be caused by Ca^{2+} -dependent K^+ channels in real neurons. Neuron k produces an instantaneous action potential when u_k reaches a threshold of 10 mV, and u_k is then reset to 0 and held at that value for 2 ms to produce an absolute refractory period. Each action potential produced by neuron k also allows a momentary burst of Ca^{2+} ions to flow into that cell (through high-potential Ca^{2+} channels) and increments J_k upward. Ca^{2+} also leaks out, with a characteristic time τ_{unadapt} usually set at 5 s. Because J_k and the internal Ca^{2+} ion concentration of neuron k are proportional, the adaptive effect can be written in terms of the variables J_k , and the cellular internal Ca^{2+} concentration is needed only to understand a possible mechanism of spike-frequency adaptation. The timescale of adaptation is set by the size of increment to J_k that occurs when a neuron spikes. The parameter I_{bias} sets the resting potential of the neurons.

When modeling a spiking network with self-sustained collective behaviors as in the present case, it is necessary to have enough exciting spikes arriving within a synaptic integration time that the self-sustained excitation is not likely to extinguish during a random fluctuation of incoming spikes. This requires the product (typical spiking rate of active place cells) \times (synaptic current duration) \times (total number of place cells) \times (fraction of place cells to which any particular cell is synaptically coupled) $\gg 1$. The speed of practical simulations in Matlab limited the total number of place cells in area E to 2000. For reasonable resolution of the place cell system, the fraction of coupled place cells was set at 0.03. To satisfy the inequality strongly, we made the firing rates of highly active neurons about 40 Hz, larger than typical in the hippocampus. Excitatory synaptic currents within area E and from area E to area A are modeled as having an instantaneous rise and an exponential fall, with a time constant of 25 ms. Because the mode of operation of the overall system can be captured by rate-based equations, the same results are expected of a system with more total neurons and smaller firing rates and synaptic time constants.

Area E implicitly contains a set of inhibitory interneurons whose function is to limit the total activity this area, which would otherwise run to very high rates due to the excitatory interconnectivity. Because this inhibitory feedback is global, and because this essential function is computationally trivial, its effect is modeled in a continuous fashion and using a single dynamical variable rather than by using spiking interneurons. A variable x represents the input current to this inhibitory system from the excitatory cells of area E. This variable obeys

$$dx/dt = -x/\tau_{\text{inhib}} + \text{positive increment from each area E spike} \quad \text{[S3]}$$

$$I_{\text{global}} \propto x - x_{\text{thresh}} \text{ if } x > x_{\text{thresh}}; \text{ otherwise } I_{\text{global}} = 0. \quad \text{[S4]}$$

x reflects the activity of area E, averaged over the time τ_{inhib} of 30 ms by Eq. S3. An inhibitory current I_{global} was introduced into each cell, computed from Eq. S4. It can be thought of as being generated by a set of inhibitory neurons with a firing threshold

x_{thresh} and which have a firing rate that increases linearly above threshold. The same kind of construction has been used in associative memory (1).

Sensory currents (when required) for each place cell are modeled as having an isotropic Gaussian form around the center of the receptive field for that cell, with the same width and strength for each neuron. When modeling multiple environments, each cell has a receptive field in each environment, assigned randomly.

The behavior of a motor control network was simulated as a single effective neuron. It received excitatory synapses of equal strength from all area A cells. Excitatory synaptic currents rise instantaneously and decay exponentially with a 40-ms time constant. The neuron also received (indirect) inhibitory synapses from all area A cells, with a 20-ms time constant, and balanced in strength so that when the sum of the firing rates of the area A cells was constant, the motor control neuron received no input. This combination, represented by I_A in Eq. S5, results in a positive current to the motor control when the total spiking rate of area A is falling. I_b is a bias current

$$du_k/dt = -u_k/\tau_{\text{cell}} + I_b + I_A - J \quad \text{[S5]}$$

$$dJ/dt = -J/\tau_{\text{recovery}}. \quad \text{[S6]}$$

J represents an inhibitory pathway, that in the single-neuron description becomes an inhibitory self-connection that has a characteristic decay time $\tau_{\text{recovery}} = 40$ ms. The motor control neuron has the same time constant and threshold for spike generation as the neurons in areas A and E. However, in operation it differs significantly from area A and E neurons in having a large increment to J when an action potential is fired, an increment so large that the cell does not fire again for many milliseconds.

When this cell fires an action potential, the direction of motion of RR reverses (velocity vector $\mathbf{V} \rightarrow -\mathbf{V}$ + directional noise), with noise distributing the resulting angle over a spread half-width of $\pm 60^\circ$. Because the inhibition from J keeps the cell from firing immediately again, the RR will keep moving in roughly this same direction for a while. By the time the cell can fire again, the input to it will reflect whether this is a good direction of motion (and the cell will not fire) or a poor direction (cell will fire). In typical operation, this cell is firing in the 5–10 Hz range. The direction of motion has Gaussian random noise introduced at each time step to mimic the effect of environmental and system noise. This produces a random walk of the direction of motion in the absence of control spikes, and is apparent in Fig. 1A.

There were 2016 neurons simulated in area E, and 1008 in area A. Because of noise in the total synaptic current into a cell due to the limited number of action potentials, the network sizes cannot be made much smaller without destroying the stability of the relevant bump attractor behavior. For example, if area E is reduced to 1008 neurons, undesirable (for the task at hand) spontaneous jumps between environments tended to occur every few seconds. Simulations were carried out using a Euler integration of the differential equations and a 0.2-ms time step. Experiments were also carried out with additive Gaussian noise, introduced at each time step to represent the effect of other sources of noise on action potential timing. Membrane noise at reasonable levels produced only small effects because the operation of the system is chiefly captured by rate-based considerations.

The rule for the spike-timing-dependent synaptic plasticity and the formation of synapses within area E and between areas E and A is described in the main text. A grid with regularly placed centers was placed on each environment, and a place cell was

positioned at the location of each of these centers plus random Gaussian noise. These locations were used for the centers of the sensory signals that drove the E and A neurons during the learning exploration. The learning exploration used a random RR motor-generated trajectory. A first attempt was made to simply put down place cells at random, but the fluctuations in place cell density led to a terrain that was too wrinkled for complete mental exploration. This problem would not occur if the number of cells was increased by a factor of 10–100, or could be reduced by additional adaptive mechanisms in a real system. The interesting possibility that two environments α, β could be designed to have the same sensory input at two locations, one in α and one in β , was ignored.

The adaptive system has a limited ability to push a bump up what would be a hill in a nonadaptive landscape. It contends best with hills when its rate of adaptation is relatively high. This is not a problem for mental exploration, which takes place rapidly when the rate of adaptation is high. It can become a problem when trying to guide a physical behavior with the very simple but relatively ineffective motor control system, for it cannot keep up with a rapid mental trajectory. In the simulations, this became an issue only in the case of Fig. 5, where the terrain had to be made quite flat in order that the mental motion not get stuck when the rate of adaptation was set low in order for the physical motion and motor control to be able to follow. Flattening was achieved by using only a single environment for Fig. 5, and increasing the exploration time for establishing intra-area-E synapses by a factor of 10 to decrease the level of noise in the structure of the network. Such a procedure was not necessary for Fig. 6 or for the movies, because the motor control system performs better in the quasi-one-dimensional situation of the T-shaped environments. This speed issue is not a fundamental limitation to the use of these mental exploration ideas in biology, where evolution can improve adaptive propulsion performance, homeostasis can smooth out a wrinkled terrain, and more effective motor control systems made use of. Indeed, the limitations of the present control system arise from noise in the spike arrivals from area A, and would be reduced to insignificant if 10 times as many neurons had been used in areas A and E.

Two other models of adaptation were tried, one based on synaptic adaptation due to vesicle depletion and one on an adapting threshold for firing. Both produced results that are quantitatively similar to those presented here. The model presented was chosen because it is easily related to an adapting energy function landscape description when spiking neurons are replaced by a rate-based description with continuous variables, and thus seems the best candidate for deeper mathematical analysis.

Although Eq. S2 was described in terms of Ca^{2+} fluxes and a Ca^{2+} -dependent inhibitory current, its form is mathematically equivalent to that of an inhibitory self-synapse with a time constant τ_{unadapt} . This can be represented by a negative diagonal term in a synaptic connectivity matrix. If a conversion is made to rate-based mathematics, and replacing I_{global} by all-to-all inhibitory synapses of strength ϵ , the entire E system is equivalent to a set of nonadaptive neurons with synaptic connections. A set of rate-based equations can be written as below. The firing rate $f(i)$ of the model neuron as a function of input current i increases monotonically from zero to a maximal value as i increases.

$$di_k/dt = -i_k/\tau_{\text{syn}} + \sum S_{kj}f(i_j + I_{\text{sensory}} + a_j) \quad \text{[S7]}$$

$$da_k/dt = -a_k/\tau_{\text{unadapt}} - \mu f(i_k + I_{\text{sensory}} + a_k) \quad \text{[S8]}$$

$$S_{kj} = T_{kj} - \epsilon \quad \text{[S9]}$$

The connection matrix S is the same as T with an additional (negative) term $-\epsilon$ representing the effect of inhibition. The

variables a_k describe the adaptation process. Because T is symmetric (or very nearly so), S is also.

The term involving μ has the mathematical form of a self-connection, leaving the overall connection pattern symmetric. Why, then, does this system have a long-term dynamical behavior that is different from that of simple symmetric networks, and that gives it fundamentally new capabilities? If $\tau_{\text{syn}} = \tau_{\text{unadapt}}$, a change of variables to $x_k = i_k + a_k$ allows the second equation to be combined with the first, yielding

$$dx_k/dt = -x_k/\tau_{\text{syn}} + \sum R_{kj}f(x_j + I_{\text{sensory}}) \quad \text{[S10]}$$

$$R_{kj} = T_{kj} - \epsilon - \mu\delta_{kj}. \quad \text{[S11]}$$

R is symmetric, and this system has the usual Lyapunov function (2). When, however, the unadaptation time constant is long compared to the synaptic time constant (the ratio used was typically 200:1) this is not the case, and exploratory behavior results. Eqs. S7–S9 describe most of the exploratory activity, although lacking the effects of “noise” caused by spiking on pathway choice and pathway persistence. (Without spike noise, the performance of the simple motor control system can be dramatically improved.)

If the adaptation parameters a_k are held fixed and there is no sensory input I_{sensory} , a Lyapunov function can be written for Eq. S7 with the notation $f_k = f(i_k + a_k)$:

$$L = 1/\tau_{\text{syn}} \int f_k^{-1}(x) dx - 1/\tau_{\text{syn}} \sum a_k f_k - 1/2 \sum \sum S_{kj} f_j f_k. \quad \text{[S12]}$$

Because the values of a_k when frozen at an instant of time during the natural evolution of the system are negative, the second term in L is positive. In the case of Fig. 3 *G* and *H*, this positive term can be decreased more rapidly by the bump moving to the right (where the bump has not yet been and the system is not now adapted) than to the left, and thus the bump will move to the right.

Activity Movies. Movies S1 and S2 show what an experiment with the quality of cellular recording now available in rodent multi-electrode studies in the hippocampus might find during one particular behavior of the RR. The behavior of the RR is appropriate to a task in which a water reward is available first at one of the three ends of a T maze, then at another (different from the previous one), then at another (different from the previous one). . . . The frame rate is 10 s^{-1} . The red asterisk shows the instantaneous position of the RR at each frame time. Each cell is displayed as a dot at the spatial location of the center of its receptive field. A dot is present or absent in a frame according to whether the cell has spiked within the previous 0.1 s. Thus, the dots represent strongly active cells. One hundred place cells (area E) or 50 place cells (area A) were chosen to typify what can currently be done in rat isolated unit multi-electrode hippocampal recording. The actual spatial trajectory between successive frames is complex, so there is a small intrinsic ambiguity in the implied comparison of instantaneous position with neural activity defined over a time interval.

The E area Movie S1 has a superficial appearance conforming to the conventional description “the animal moves from location to location (cause of that motion not specified), and the pattern of neurons which are active follow that motion because they are indirectly coupled to sensory stimuli which drive them, and have location-specific sensory receptive fields.”

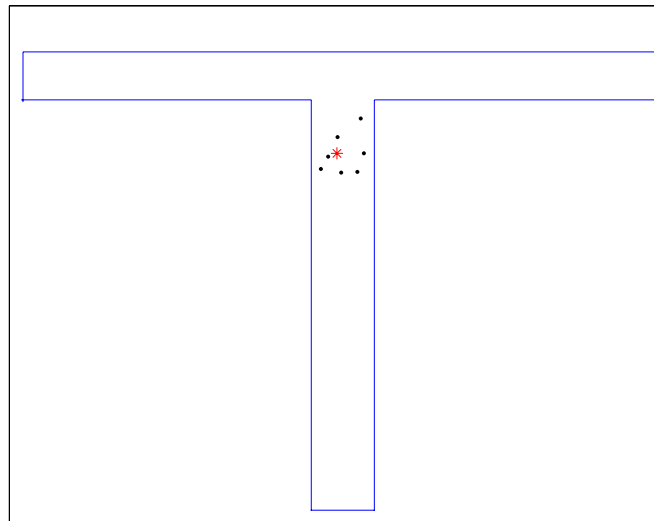
Yet there is sensory input to area E only in the first frame of the movie, setting an initial mental state that corresponds to the initial location of the RR. In all other frames there is no sensory input to E. The correct explanation of the movie is: “The animal is located near the mean of the place cell centers of the strongly active neurons because there is a motor control system that causes the

animal to make its physical location correspond to a bump activity pattern present in E. The activity pattern in area E is a self-generated bump that moves around spontaneously, *with no sensory input whatsoever to E.* The appropriate metabehavior of moving from one limb of the T to another is spontaneously generated by area E. This becomes the metacommand that (using a comparison with real sensory signals done in area A) results in actual motor commands which cause the position of the RR in real space to correspond to the bump location. The cause and effect that are responsible for the correlation between activity bump and physical location in the movie are reversed from that of the conventional description of the previous paragraph.

Movie S2 of area A place cell activity is made in the same fashion. These cells have both sensory inputs and inputs from area E. The movie is superficially similar to **Movie S1**. Careful inspection suggests that the fraction of active cells can vary more in area A, but the effect is subtle when so few cells are recorded from. The movies also suggest the presence of a little “anticipation” on the part of the place cells, especially of area E. This is, of course, not anticipation in terms of sensory system anticipation (which in a real animal might be done on the basis of the statistics of previously experienced dynamical stimuli), but is instead due to the fact that causative neural activity must precede motor actions.

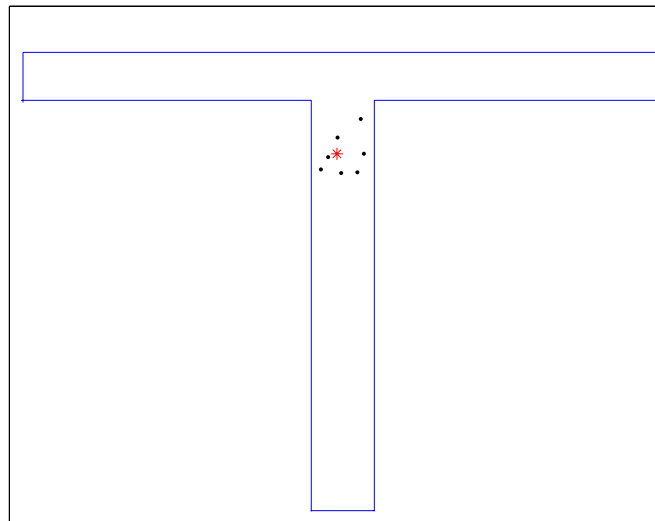
1. Hopfield JJ (2008) Searching for memories, Sudoku, implicit check bits, and the iterative use of not-always-correct rapid neural computation. *Neural Comput* 20:1119–1164.

2. Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. *Proc Natl Acad Sci USA* 81:3088–3092.



Movie S1. RR position and place cell activity in area E during behavior.

[Movie S1](#)



Movie S2. RR position and place cell activity in area A during behavior.

[Movie S2](#)