A Closer Look at Classical Measurement, an Algorithm for Deliberation in Rodents, and a Conjecture on Intertemporal Choice

by

David Francisco Theurel

Lic., Univ. Nacional Autónoma de México (2013)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Author .....
Department of Physics
May 13, 2022

Certified by .....
Matthew A. Wilson
Professor of Neuroscience
Thesis Supervisor

Certified by .....
Mehran Kardar
Professor of Physics
Thesis Supervisor

Accepted by .....
Deepto Chakrabarty
Associate Department Head of Physics

A Closer Look at Classical Measurement, an Algorithm for Deliberation in Rodents, and a Conjecture on Intertemporal Choice

by

David Francisco Theurel

Submitted to the Department of Physics on May 13th, 2022
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Physics

Abstract

In this three-part thesis, Part I is an examination of the measurement process in classical Hamiltonian mechanics. This part is concerned with the tradeoff that exists, when measuring any observable of a system, between the disturbance inflicted upon the system and the information that can be extracted. The main result takes the form of a Heisenberg-like precision-disturbance relation: measuring an observable leaves all compatible observables undisturbed but inevitably disturbs all incompatible observables. The magnitude of the disturbance (the analogue of $\hbar$) is found to be proportional, in a sense that is made precise, to one's initial uncertainty in the ready-state of the apparatus—a quantity that relates to the temperature of the apparatus.

Part II of this thesis develops a model of the computations taking place in the deliberative decision-making system of rodents, during wakefulness and sleep, with focus on the role of hippocampus (HPC). In this model, medial prefrontal cortex performs high-level planning, and then tasks HPC with fleshing out the details of the plan, as needed. We describe this planning task of HPC as an optimal control problem, which allows us to draw insights from the powerful mathematics of optimal control theory. The model makes novel testable predictions, provides insights into memory consolidation during sleep, and offers a paradigm capable of accommodating a wide range of observed phenomena, such as the theta rhythm, the slow oscillation, spindle oscillations, sharp wave-ripples, θ-sequences, forward and reverse SWR-sequences, the formation and strengthening of episodic memories, and a need for two modes of operation—online and offline.

The two parts described above are the main content of this thesis. Part I falls within the purview of classical theoretical physics, while Part II falls in that of computational neuroscience. The two may seem unrelated; however, while each part is self-contained, I see the two as connected. Part III of this thesis is my attempt to provide an outline of a bigger picture, which sees the foregoing as lines of inquiry towards the same far-reaching conjecture—one which has had a strong pull on my imagination during my PhD, and which I hope to be able to address in the future. This conjecture is that the probability calculus of quantum mechanics holds a kind of normative status for a class of decision problems involving intertemporal choice under uncertainty—a class of problems of great importance to artificial intelligence, brain sciences, economics, and, I argue, to physics too.

Thesis Supervisor: Matthew A. Wilson
Title: Professor of Neuroscience

Thesis Supervisor: Mehran Kardar
Title: Professor of Physics

Acknowledgements

I'd like to express my gratitude to the Dean of Science Fellowship Program for financial support during the first half of my PhD; as well as to the Department of Physics and the England lab, for financial support during an intermediate transitional period of my PhD. Likewise, I'm grateful to the Picower Neurological Disorder Research Fund, and in particular the Wilson and Brown labs, for financial support during the second half of my PhD. The freedom all of you have granted me has allowed me to find my way. Thanks to Cathy Modica for saving my butt every time I got into trouble with the Physics Department.

It is a pleasure to thank Eulogio, Pavel and Kurt Jacobs for reading the material that makes up Part I of this thesis, and providing many useful suggestions. Many thanks to Sage, my collaborator for the material of Part II, for your resourceful ideas and for being so very patient with me.

I thank my friend Pavel, and my old friends 'la banda de la fac' from UNAM, for always fostering my love of physics. I thank my even older friends 'la palomilla' from La Paz, for always reminding me to laugh and be lighthearted. Without all of you I would've lost my mind over quarantine. Love and gratitude to my family back home, specially my parents. (Now will you stop teasing me about giving me a job flipping eggs at the B&B, if I don't make it as a scientist?!)

Thank you Tai, for your love, support, patience and good advice.

To the members of the Wilson lab past and present—Carmen, Tim, Hannah, CJ, Maz, Alexis, Jon, Hector, Pedro, Honi, Jack, Qianli, Francisco, Takato, Wei, Isabella and Matt—your tireless hard work and wild creativity have been a source of wonder and inspiration. Our many discussions and happy hours have made all the difference during my time at MIT.

I'm at a loss for words to thank Matt, for believing in me when I most needed it, for giving me a home while I found my way, and for your continued mentorship and patience. I wouldn't have made it to this point of my PhD without your help.

Contents

Part I

A closer look at classical measurement[1]

Abstract

Measurement in classical physics is examined here as a process involving the joint evolution of an object-system and a finite-temperature measuring apparatus. For this, a model of measurement is proposed which lends itself to theoretical analysis using Hamiltonian mechanics and Bayesian probability. At odds with a widely-held intuition, we find that the ideal measurement capable of extracting finite information without disturbing the system is ruled out. In its place we find a Heisenberg-like precision-disturbance relation: measuring an observable leaves all compatible observables undisturbed but inevitably disturbs all incompatible observables. In this classical uncertainty relation the role of $\hbar$ is played by an apparatus-specific quantity, q. While this is not a universal constant, our model suggests that q takes a finite positive value for any apparatus that can be built. (Specifically: q vanishes in our model only in the unreachable limit of zero absolute temperature.) Additionally, the process of continuous measurement is examined, yielding a novel Liouville-like master equation describing the dynamics of (a rational agent's knowledge of) a system under continuous measurement. The resulting equation is analogous to the stochastic master equation used in continuous quantum measurement. I believe the approach presented here points the way to studying the (Bayesian) epistemology of classical physics, which has until now been overlooked and wrongly assumed to be trivial. These results suggest that said epistemology has instead a non-trivial structure bearing a resemblance to the quantum formalism. For this reason, these findings may be of interest to researchers working on the foundations of quantum mechanics, particularly for ψ-epistemic interpretations. More practically, these results may find applications in the fields of precision measurement, nanoengineering and molecular machines.

I.1 Introduction

It is commonly held among the wider physics community that the topic of classical measurement is essentially trivial. I don't mean the modeling in physical detail of any one laboratory setup, which of course can get very complicated, but just the examination of "measurement" as a bare-bones physical process, idealized away from as many complications as possible; a theoretical physicist's model of measurement. One way of stating the wide-held intuition is that there is in principle no obstruction in classical physics to measuring any observable of a system with arbitrary precision while disturbing the system arbitrarily little. This intuition is in sharp contrast to the situation in quantum physics, where the Heisenberg uncertainty principle (specifically the Ozawa inequality [1]) asserts just such a limit. Surely influenced by this attitude, there is a correspondingly sharp

[1]This Part is adapted from my paper arXiv:2104.02064 [quant-ph], submitted for publication to the journal Physical Review E.

contrast between the little attention ever payed to the measurement process in classical physics, and the large attention payed over the decades (deservedly) to that same process in quantum physics. To the best of my knowledge, only a handful of examples can be attributed to the first category: Heisenberg's own thought experiments in the late 1920's [2] (particularly Heisenberg's microscope); although they served as the motivation for his quantum uncertainty principle, they were essentially classical arguments, augmented only by Einstein's theory of the photon. In 1996 Lamb and Fearn [3] set up the problem of a classical point particle (the system) in interaction with a second point particle (the "apparatus") subject to noise. They stopped short of a thorough analysis; their primary interest being the quantum case. Recently Morgan [4] and Katagiri [5] made use of KvN formalism in independent attempts to use quantum measurement theory to examine measurement in classical mechanics.

The only long-lasting foray into classical measurement seems to be within the body of work surrounding Maxwell's demon and the foundations of thermodynamics. The demon was first conceptualized by Maxwell in 1867 [6] as a "very observant and neat-fingered being" capable of monitoring the molecules of a gas, and, by opening and closing a small door without exerting any work, of sorting the high-energy molecules from the low, thus creating a temperature gradient. This amplifier of fluctuations, if it existed, could then be used to run a perpetual motion machine of the second kind, violating the second law. Writing in 1929 Szilárd [7] realized that, if the second law was to hold, somewhere in the demon's monitoring of the molecules (i.e. in the measurement process) entropy had to be produced. Soon afterwards von Neumann [8], in his reading of Szilárd, pointed to information acquisition as the key step incurring entropy cost. The latter claim was developed prominently in the 1950's by Brillouin [9, 10] and Gabor [11]. But in the 1980's Bennett [12, 13] (building on work by Landauer [14]) argued against Brillouin and Gabor, pointing instead to erasure of the measurement record as the key step incurring entropy cost. This 150-year-long inquiry may be finally nearing a close in recent years, with the answer appearing to be that both sides, Brillouin-Gabor and Bennett, had part of the answer: and that the entropy cost of measurement can be traded between the acquisition and erasure steps. This resolution is reviewed in [15], in an analysis that relies on quantum (not classical) measurement theory.

The above illustrates three points which I would like to contend: (i) despite the wide-held intuition, measurement in classical physics is far from trivial; (ii) it is a surprisingly underdeveloped subject; and (iii) unacknowledged, it is a subject whose immaturity may have long held back progress in some fields of physics. To address the issue, a reasonable aim would be a theory of measurement in the context of Hamiltonian mechanics, which is the mathematical framework at the foundation of classical physics. The research program I'm suggesting can be summarized as: to systematically bring Bayesian probability to bear on an ontology governed by classical Hamiltonian mechanics, with the full strength, and no more, that is permitted by the geometro-algebraic structure of the ontology. That is; to develop the (Bayesian) epistemology of classical Hamiltonian ontology. The present paper aims to kickstart this program, with no ambition of being the final word.

We begin by noting that the assumption of perfect information regarding the initial state of the measuring apparatus is unrealistic. In fact it is ruled out as a matter of principle by the third law of thermodynamics; initial uncertainty must be present if for nothing other than for finite-temperature thermal noise. Next we posit a model of the measurement as a physical process. While some minimal assumptions are made concerning the systems that can be used as measuring apparatuses, no restrictions are placed on the system under measurement. This model enjoys substantial generality while at the same time lending itself to Bayesian analysis. We then show that, in the process of measurement, the uncertainty in the state of the apparatus propagates into two uncertainties regarding

the object-system: one is the imprecision of the measurement; and the other an uncertainty in the magnitude of the disturbance caused upon the system—that is, an observer effect. And we find that these two are bound by a Heisenberg-like precision-disturbance relation. In particular, while we find no obstacle in principle to making a measurement arbitrarily precise, we do find an obstruction to realizing such a measurement without disturbance. Thus our findings are at odds with the wide-held intuition.

Interestingly, the disturbance in question is not arbitrary but takes the particular form of time-evolution under the Hamiltonian flow generated by the measured observable; the only thing uncertain is how much "time" the system flowed. Thus observables in involution (i.e. "compatible") with the one measured are spared, while those not in involution (i.e. "incompatible") are disturbed. A general consequence of these results seems to be that in classical physics, like in quantum physics, observables can be simultaneously perfectly-precisely measured if and only if they are compatible.

Next, we derive a novel Liouville-like master equation describing the dynamics of (a rational agent's knowledge of) a system under continuous measurement. This equation, which is analogous to the stochastic master equation appearing in continuous quantum measurement [16], is capable of describing general sequences of measurements, including inefficient measurements and simultaneous measurements of multiple observables.

While I hope our topic will be of interest to several fields of physics, it may be of particular interest to ψ-epistemic interpretations of quantum mechanics; given that we find indications that the epistemology of classical Hamiltonian mechanics is more similar to the quantum formalism than has previously been recognized.

The rest of the Part is organized as follows. We begin in Section I.2 by reminding the reader of the basic concepts and equations of Hamiltonian mechanics. Section I.3 does the conceptual heavy lifting; there we construct our measurement model and obtain the basic results on which the rest of the paper is based. In Section I.4 we arrive at the precision-disturbance relation. In Section I.5 we consider the problem of continuous weak measurement over time, which enables us to discuss simultaneous measurements of multiple observables, as well as inefficient measurements. The method of analysis there is drawn directly from the field of continuous quantum measurement. In Section I.6 we discuss a few relevant topics in the new light of our results: the similarities, and likely coexistence in the real world, of the classical and quantum uncertainty relations; the epistemic limitations inherent to classical Hamiltonian ontology; and the subtle interplay between ontology and epistemology in a theory of measurement. The paper ends by contemplating some of the many possibilities ahead; both the concrete and the speculative.

I.2 Brief recap of Hamiltonian mechanics

Hamiltonian mechanics is a confluence of differential, algebraic and symplectic geometry, Lie algebra and Lie groups. A wonderful resource for the topic is [17].

We consider a continuous-time dynamical system over a $2n$-dimensional symplectic manifold, called *phase space*. The *observables* of the system (e.g. position, momentum, angular momentum, etc) are the smooth, single-valued, real-valued functions defined globally over phase space. By convention we take observables to not depend explicitly on time. (With this convention, any explicit time-dependence is regarded as specifying a different observable at each moment in time.) The points in phase space can be expressed in local *canonical coordinates* $(q,p) = (q_1,\ldots,q_n,p_1,\ldots,p_n)$ (Darboux's theorem). In terms of these coordinates, the state of the system evolves over time according to *Hamilton's equations*,

$$\dot{q}(t) = \frac{\partial H}{\partial p}(q(t),p(t);t), \quad \dot{p}(t) = -\frac{\partial H}{\partial q}(q(t),p(t);t), \tag{1}$$

where at each moment the system's *Hamiltonian*, H, is an observable. Notice that "Hamiltonian" and "H" are indexical terms; they don't specify any concrete function over phase space, but refer to whichever observable happens to serve as the generator of time-evolution (as in (1)) for a given system at a given time. At each moment Hamilton's equations describe a *flow* Φ_τ^H on phase space. Along the integral curves of this flow the value of any observable $A(q,p)$ changes as

$$\dot{A} = \{A, H\}, \tag{2}$$

where $\{A, H\}$ denotes the *Poisson bracket*,

$$\{A, H\} \triangleq \sum_{j=1}^{n} \left(\frac{\partial A}{\partial q_j}\frac{\partial H}{\partial p_j} - \frac{\partial A}{\partial p_j}\frac{\partial H}{\partial q_j} \right). \tag{3}$$

(Note that (2) follows from (1) after application of the chain rule to $\frac{d}{dt}A(q,p)$; but also contains (1) as special cases when A equals one of the canonical coordinates.) Two observables A, B for which $\{A, B\}$ is identically zero are said to be in *involution* with each other. In this case, by (2), the value of A remains constant along the integral curves of the flow Φ_t^B (and vice versa). It follows that any observable in involution with the Hamiltonian is a *constant of the motion*. In particular, if H is not explicitly time-dependent then it is itself a constant of the motion (conservation of energy). Including itself, a given observable can be in involution with as few as one and as many as $2n$ independent observables, but only as many as n independent observables can be all in involution with one another. On the other hand, if $\{A, B\} = 1$ identically then A, B are said to be *conjugate* to each other. In this case B is also said to be "the" *generator of translations* in A (and vice versa); because, by (2), the value of A changes monotonically at unit rate along the integral curves of the flow Φ_t^B. A given observable, A, may fail to have a conjugate observable. In this case, in a neighborhood of any regular point of A (i.e. where $dA \neq 0$), it is still possible to speak of a locally-defined conjugate "quantity", B, which satisfies $\{A, B\} = 1$ but fails to satisfy the stringent definition of a bona fide observable. This is illustrated on the 2D phase space by the observable $I = \frac{1}{2}(q^2 + p^2)$ (the Hamiltonian for the simple harmonic oscillator); whose conjugate quantity $\phi = \arg(q + ip)$ (the phase of oscillation for the sho) either fails to be globally continuous, or else fails to be single-valued (depending on one's choice of definition).

Notice that the components of (q,p) satisfy the *canonical relations*

$$\{q_i, q_j\} = \{p_i, p_j\} = 0, \quad \{q_i, p_j\} = \delta_{ij}, \tag{4}$$

so each canonical coordinate is in involution with all other coordinates but one, to which it is conjugate. A diffeomorphism of phase space, $(q,p) \mapsto (q',p')$, such that (q',p') again satisfy these canonical relations is said to be a *canonical transformation*. Canonical transformations have Jacobian determinant equal to 1, so they preserve the *Liouville measure* of phase space volume, $d^n q d^n p = d^n q' d^n p'$. For any flow parameter, τ, the Hamiltonian flow Φ_τ^H is an example of an (active) canonical transformation; in particular, Hamiltonian flow preserves the Liouville measure (Liouville's theorem). Changes of coordinates implemented by (passive) canonical transformations are particularly convenient since they preserve the simple form of the Liouville measure, the equations of motion (1, 2), and the Poisson bracket (3).

I.3 A model of measurement in a Hamiltonian world

Suppose we wished to measure an observable $A(q,p)$ of the system (1) at time t_0. In the world of Hamiltonian mechanics this can only be done by coupling the system to a

measuring apparatus, where the joint system (= object-system + apparatus) is itself a Hamiltonian system, with

$$H_{\text{joint}}(q,p,x,y;t) = H(q,p;t) + H_{\text{app}}(x,y;t) + H_{\text{int}}(q,p,x,y;t). \tag{5}$$

Here (x,y) are canonical coordinates on the $2m$-dimensional phase space of the apparatus; H_{app} is the apparatus' Hamiltonian; and H_{int} is the interaction between system and apparatus, which we will assume to be switched on only briefly around $t = t_0$. We now stipulate a model for the measurement.

I.3.1 System-apparatus coupling

Consider the "gauge", or "pointer display", of the apparatus; by which I mean the observable of the apparatus which, after interaction with the system, we want to reflect the sought-after value of A at time t_0. Denote this observable of the apparatus by $P(x,y)$. Suppose P has a conjugate observable, $Q(x,y)$ (so that $\{Q,P\} = 1$). For the interaction to imprint the value of A on P, the interaction Hamiltonian must involve A and the conjugate quantity to P, namely Q; because this is the generator of translations in P.[2] The simplest interaction of this form is the product[3]

$$H_{\text{int}}(q,p,x,y;t) = \alpha\delta(t-t_0)A(q,p)Q(x,y), \tag{6}$$

where $\alpha \in \mathbb{R}\backslash\{0\}$ is a constant of proportionality, and $\delta(t-t_0)$ is the Dirac delta function indicating that the measurement is idealized as taking place instantaneously at t_0.

I.3.2 Readying the apparatus

Let us take a step back to consider how to initialize the apparatus into its "ready state" prior to interaction at t_0. Being, as we are, in the process of defining what we mean by "measurement", on pain of circularity we shouldn't appeal to measurement to assess the state of the apparatus, as might be needed to actively manipulate it into a state ready for measurement of the system. This difficulty can be circumvented by letting low-temperature thermalization take care of confining the state of the apparatus to a narrow region of its phase space. The region in question can be specified experimentally by setting up a deep energetic well there—a "trap". This trap could be due to a confining gravitational or electrostatic potential; a combination of near-field electric and magnetic fields; a light field; atomic chemical bonds; etc. We write

$$H_{\text{app}}(x,y;t) = H^{\text{own}}_{\text{app}}(x,y) + \Pi(t)H_{\text{trap}}(x,y), \tag{7}$$

where $H^{\text{own}}_{\text{app}}$ is the apparatus' own, or internal, Hamiltonian, which we take to be time-independent; and $\Pi(t)$ is a rectangular step-function taking only the values $1/0$, describing the on/off switch of the trap. The trap will be switched off for all $t > t_0$; it is only switched on in the time leading up to t_0, to help bring the apparatus into its ready state, as we

[2] To be more precise: for any specified pointer $P(x,y)$, by the Carathéodory-Jacobi-Lie theorem [18] there exists, in a neighborhood of any regular point of P (i.e. where $dP \neq 0$), a canonical coordinate system for the apparatus in which P is one of the coordinates. By Q we mean the coordinate conjugate to P in this system. The requirement that Q be a bona fide observable amounts to the non-trivial assumption that this coordinate can be extended to a smooth single-valued function globally on phase space. As seen in (4), P is in involution with all other coordinates of this system but Q. It follows that if, upon expressing H_{int} in these coordinates, Q did not appear, then we would have $\{H_{\text{int}}, P\} = 0$; and by (2) the interaction would have no immediate effect on the pointer P. Since this is the opposite of what we want, we see that H_{int} should depend on Q.

[3] Note that this interaction is the classical analogue of that used in the von Neumann model of quantum measurement [8].

will now describe. The trap consists of a deep energetic well which, when switched on ($\Pi(t) = 1$), sets the ground state of the apparatus at some point (x^*, y^*) of its phase space. Without loss of generality we may set our coordinates such that $(x^*, y^*) = (0, 0)$, and we may assume that the corresponding energy is $H_{\text{app}}(x^*, y^*)|_{\text{trap on}} = 0$. (Otherwise these conditions can be met by shifted redefinitions of x, y, H_{app}.) We Taylor-expand $H_{\text{app}}(x, y)|_{\text{trap on}}$ around the ground state, obtaining a positive-definite quadratic form:

$$H_{\text{app}}(x, y)\Big|_{\text{trap on}} = H_{\text{app}}^{\text{own}}(x, y) + H_{\text{trap}}(x, y) = \frac{1}{2}\begin{pmatrix} x & y \end{pmatrix} \hat{M} \begin{pmatrix} x \\ y \end{pmatrix} + \mathcal{O}(3), \tag{8}$$

where $\hat{M}$ is a symmetric positive-definite $2m$-by-$2m$ matrix of coefficients, and $\mathcal{O}(3)$ denotes all higher-degree terms in the series. As shown by Whittaker [19] (see also theorem by Williamson [20], explained in [17, appendix 6]), there exists a local linear canonical coordinate transformation $(x, y) \mapsto (z, w)$ which reduces (8) to the normal form

$$H_{\text{app}}(z, w)\Big|_{\text{trap on}} = \frac{1}{2}\sum_{i=1}^{m}(b_i^2 z_i^2 + w_i^2) + \mathcal{O}(3). \tag{9}$$

Here $b_1 \geq b_2 \geq \cdots \geq b_m > 0$ are constants with physical dimensions of angular frequency; they are the natural frequencies of oscillation of the apparatus around its trapped ground state.

Now to ready the apparatus: while the trap is on, the apparatus is brought into contact with a thermal bath at some temperature $T = 1/\beta k_B$, allowed to equilibrate, and then isolated again.[4] After this our knowledge about the state of the apparatus is given by the Boltzmann probability distribution

$$\rho(z, w)\, d^m z d^m w \propto \exp\left\{-\beta H_{\text{app}}(z, w)\Big|_{\text{trap on}}\right\} d^m z d^m w. \tag{10}$$

Note that in the time between isolation from the bath and measurement at t_0 the evolution of the apparatus will preserve this distribution, as opposed to spoiling the preparation, since $H_{\text{app}}|_{\text{trap on}}$ is constant under the phase-space flow generated by itself and such flow preserves the Liouville measure $d^m z d^m w$.

At this point we make three requirements that constrain the apparatuses, traps, and temperatures allowed by our model. (i) We require that the trap be harmonic enough, or the temperature be low enough, that in the Boltzmann distribution (10) the higher-degree terms in (9) can be neglected. (ii) We require that at least one of the coordinates w_i be in involution with $H_{\text{app}}^{\text{own}}$. Let $i = i^*$ be the index of this special coordinate. (If given a choice, we want the associated frequency b_{i^*} to be as large as possible, for a reason to be seen in Section I.4.) The condition means that w_{i^*} will be a constant of the motion of the apparatus when the trap is switched off—a desirable property for the pointer P (introduced in Section I.3.1); so that the measurement record is stable after the interaction has past. We thus identify the pointer $P \triangleq w_{i^*}$ and its conjugate $Q \triangleq z_{i^*}$. We denote the corresponding frequency by $\Omega \triangleq b_{i^*}$. Note the physical interpretation of Ω as the natural frequency of oscillation of the pointer around its trapped state. Since Q, P are required to be observables, in making these identifications we're implicitly making our assumption (iii): the pair of conjugate local quantities (z_{i^*}, w_{i^*}) are globally extendable to smooth single-valued functions on phase space.

From now on Q, P are the only observables of the apparatus with which we will be concerned. With the above requirements met, we can easily marginalize over all other

[4] Instead of removing the bath, we might require just that its coupling to the apparatus be weak enough that it doesn't spoil the measurement record, P, on the timescales of interest.

variables in (10) to find the probability distribution over the pointer and its conjugate:

$$\rho(Q,P)dQdP = \frac{\beta\Omega}{2\pi}\exp\left\{-\frac{\beta\Omega^2}{2}Q^2 - \frac{\beta}{2}P^2\right\}dQ\,dP. \tag{11}$$

This is the apparatus ready state. It describes a preparation in which the pointer and its conjugate have been set independently to zero, but there remains some uncertainty on their exact values.

I.3.3 Integrating Hamilton's equations

Integrating Hamilton's equations for the joint system, the effect of the interaction (6) is to instantaneously change the state of both object-system and apparatus as[5]

$$\begin{pmatrix} q \\ p \end{pmatrix}_{t_0^+} = \Phi^A_{\alpha Q}\begin{pmatrix} q \\ p \end{pmatrix}_{t_0^-} \tag{12a}$$

$$\begin{pmatrix} Q \\ P \end{pmatrix}_{t_0^+} = \begin{pmatrix} Q \\ P - \alpha A(q,p) \end{pmatrix}_{t_0^-} \tag{12b}$$

where Φ^A_τ is the transformation on the system's phase space that implements flowing for a "time" τ under the Hamiltonian flow generated by A. Having initialized the apparatus to its ready state (11) prior to the interaction, then, in view of (12b), after the interaction our state of knowledge of the apparatus, conditional on a given state of the system at the time of measurement, is

$$\rho(Q,P|q,p)dQdP = \frac{\beta\Omega}{2\pi}\exp\left\{-\frac{\beta\Omega^2}{2}Q^2 - \frac{\beta}{2}\left(P + \alpha A(q,p)\right)^2\right\}dQdP. \tag{13}$$

Note that the dependence on (q,p) is only through $A(q,p)$.

The trap on the apparatus is released at the moment of measurement ($\Pi(t) = 0$ for $t > t_0$), so that the apparatus Hamiltonian returns to its internal setting $H^{\text{own}}_{\text{app}}$. By construction the pointer P is in involution with this Hamiltonian, so it constitutes a stable record of the measurement. At this time (i.e. any time after t_0) we read the pointer on the apparatus, yielding some definite value P^*, or equivalently

$$A^* \triangleq -\frac{P^*}{\alpha}. \tag{14}$$

(A^* is just the reading on the pointer with the scale set appropriately.) Note that this does not mean that the value of A at the time of measurement is A^*! Rather, given this datum, the likelihood function for the value of A at the time of measurement is, from (13),

$$\rho(A^*|A)dA^* = \sqrt{\frac{\alpha^2\beta}{2\pi}}\exp\left\{-\frac{\alpha^2\beta}{2}\left(A^* - A\right)^2\right\}dA^*. \tag{15}$$

This completes our model of measurement. The *measurement record* A^*, or equivalently the likelihood function (15) (with A^* specified), constitutes the outcome of the measurement.

[5] To do this calculation it helps to approximate the δ by a square impulse of width Δt and height $1/\Delta t$. As Δt is taken smaller and smaller, the joint Hamiltonian (5) becomes dominated by H_{int} during the interaction, so that H and H_{app} can be neglected during the brief time Δt. Noting that both A and Q are constant under the flow generated by the interaction Hamiltonian (6), both parts of (12) then follow readily.

I.3.4 Consuming the measurement

There are two operations that one, as a recipient, should perform to consume the information of the measurement. The first is triggered by the information that the observable A of the system was measured at time t_0 by the stipulated procedure, with specified settings (α, β, Ω). As seen in (12a), the interaction involved in this measurement affects the state of the system by causing it to move along the flow generated by A for some unknown "time" αQ. If one knew the value of Q then one should change their probability distribution about the state of the system at time t_0 according to

$$\rho(q,p;t_0^-) \mapsto \rho(q,p;t_0^+) = \left[\left(\Phi^A_{\alpha Q}\right)_* \rho\right](q,p;t_0^-),$$

where $(\Phi^A_\tau)_*$ denotes the push-forward of the transformation Φ^A_τ, defined as $[(\Phi^A_\tau)_*\rho](q,p) \triangleq \rho(\Phi^A_{-\tau}(q,p))$; and the $+/-$ superscripts on t_0 are meant as a reminder that this update reflects a physical transition of the system that took place in a short time interval around t_0. But one does not know the value of Q; all that is know about it is expressed by the probability distribution (11). One folds this in by marginalizing over Q:

$$\rho(q,p;t_0^+) = \sqrt{\frac{\beta\Omega^2}{2\pi}} \int_{-\infty}^{\infty} dQ \, \exp\left\{-\frac{\beta\Omega^2}{2}Q^2\right\} \left[\left(\Phi^A_{\alpha Q}\right)_* \rho\right](q,p;t_0^-). \tag{16a}$$

The second operation is triggered by the information of the measurement outcome (15). One assimilates this by performing the Bayesian update $\rho_{\text{pri}}(q,p;t_0) \mapsto \rho_{\text{post}}(q,p;t_0)$, with

$$\begin{aligned}\rho_{\text{post}}(q,p;t_0) &\propto \rho_{\text{pri}}(q,p;t_0)\rho(A^*|A(q,p)) \\ &\propto \rho_{\text{pri}}(q,p;t_0) \exp\left\{-\frac{\alpha^2\beta}{2}\left(A^* - A(q,p)\right)^2\right\},\end{aligned} \tag{16b}$$

where the omitted factor of proportionality is just the normalization, obtained by integrating the expression shown over the system's phase space ($\int d^n q d^n p$). Since multiplication by a function of A commutes with the push-forward $\left(\Phi^A_\tau\right)_*$, operations (16a, 16b) can be performed in either order to the same effect. If (16b) is performed first, it corresponds to updating one's knowledge about the state the system was in before the measurement was made (i.e. at t_0^-); if second, about the state the system was left in by the measurement. Notice that if only the fact of the measurement is revealed but not the outcome (in this case we say the outcome was *discarded*), then one should only perform operation (16a), not (16b).

Finally, if a single number is desired as an objective quantification of the measured observable (i.e. not biased by anyone's prior), the maximum-likelihood estimate can be given, from (15):

$$A\Big|_{t_0} = A^* \pm \frac{1}{\sqrt{\alpha^2\beta}} \tag{17}$$

(mean ± standard deviation).[6] We will refer to

$$\epsilon_A \triangleq \frac{1}{\sqrt{\alpha^2\beta}} \tag{18}$$

as the *imprecision* of the measurement. (But notice that to translate this to an uncertainty in a given agent's knowledge of A we must first combine the likelihood function with the agent's prior, as in (16b).)

[6] The justification for calling this number a "mean ± standard deviation" is that that is what it corresponds to in the posterior (16b) when the marginalized prior $\rho_{\text{pri}}(A';t_0) \triangleq \int d^n q d^n p \, \delta(A' - A(q,p))\rho_{\text{pri}}(q,p;t_0)$ is sufficiently flat.

I.3.5 Measurement strength and apparatus quality parametrize our model

Of the three parameters (α, β, Ω) entering our model—respectively the constant of proportionality in the interaction Hamiltonian (6), the (inverse) temperature of the apparatus, and the frequency of oscillation of the apparatus' pointer around its trapped ground state—only the two combinations

$$k \triangleq \frac{\alpha^2 \beta}{8} > 0 \quad \text{and} \quad \hbar \triangleq \frac{2}{\beta \Omega} > 0 \tag{19}$$

appear independently in the final results, (16a, 16b), which can be written as

$$\rho(q, p; t_0^+) = \int_{-\infty}^{\infty} \frac{d\tau}{\sqrt{4\pi k \hbar^2}} \exp\left\{-\frac{1}{4k\hbar^2}\tau^2\right\} \left[\left(\Phi_\tau^A\right)_* \rho\right](q, p; t_0^-), \tag{20a}$$

$$\rho_{\text{post}}(q, p; t_0) \propto \rho_{\text{pri}}(q, p; t_0) \exp\left\{-4k \left(A^* - A(q, p)\right)^2\right\}. \tag{20b}$$

We will refer to k as the *strength* of the measurement; indeed, in view of (18), the larger k the higher the measurement's precision.[7] Its physical dimensions are $[k] = [A]^{-2}$.

For a reason to be seen next, we will refer to $\hbar$ ("q-bar") as the *inverse quality* of the apparatus (i.e. lower values of $\hbar$ will correspond to higher-quality devices). Note that it has physical dimensions of action.

I.4 A Heisenberg-like precision-disturbance relation in Hamiltonian mechanics

Our measurement model is characterized by the pair $(k, \hbar)$; respectively the strength of the measurement, and the (inverse) quality of the apparatus. As has just been said, we can make our measurement of A more precise by cranking up the strength, k, which we might think of as a knob on our experimental setup. However, notice that the more precisely A is measured (i.e. the larger k), the more uncertain we are about the magnitude of the back-action, or observer effect, in (20a) (i.e. the larger the variance, $2k\hbar^2$, in the "flow time", τ). It's worth emphasizing that this disturbance of the system is not arbitrary, but has the form of time-evolution along the Hamiltonian flow generated by the measured observable, A; the only thing uncertain is how much "time" the system flowed. We can see that this disturbance will affect some observables of the system more than others: in particular, any observable B in involution with A will emerge undisturbed in the immediate aftermath of the measurement (although subsequent time-evolution under the system's own dynamics will cause the initial disturbance to "leak into" such a B, unless B is also in involution with H).

Concretely, we find that the imprecision of a measurement (18), and the magnitude of the *disturbance* caused by the measurement upon the system,

$$\eta_A \triangleq \sqrt{2k\hbar^2}, \tag{21}$$

obey the inverse relation

$$\epsilon_A \eta_A = \frac{\hbar}{2}, \tag{22}$$

[7] Given the definition of A^* in (14), one might worry that not just ϵ_A, but also A^* scales with $1/\sqrt{k}$, but that's not the case: notice, from (13), that P^* is drawn from a gaussian centered at $-\alpha A$. Obviously the value of A is independent of our decision to measure it, and a fortiori of our setting of α. Hence it's the reading on the dial, P^*, that scales with α, so A^* is unaffected (in expectation) by the strength of the measurement.

which is fixed for a given apparatus quality; independent of the identity of the system measured, of that of the system used as measuring apparatus, of the measurement strength and of the choice of observable measured. The product on the left-hand side can easily be made larger (see discussion in Section I.6.1) but not smaller, as far as I can tell. To the extent that our model of measurement has a claim to generality, relation (22) will be a general principle. This Heisenberg-like precision-disturbance relation (or "uncertainty relation" for short) suggests an obstruction to how close we can come in a world governed by Hamiltonian mechanics to the idealization of measurement without disturbance. Note that this relation is softer than the Heisenberg uncertainty principle of quantum mechanics: for any given apparatus one will have a finite obstruction on the right-hand side of (22), but one can always endeavor to make the obstruction smaller by cooling the apparatus further or tightening the trap (i.e. improving apparatus quality). Instead this obstruction is of a kind with the third law, to which it is clearly related: it suggests that it is impossible by any procedure, no matter how idealized, to reduce the observer effect of measurement to zero in a finite number of operations.

I.5 Continuous measurement over time, and simultaneous measurement of multiple observables

Extracting information about the system by measurement increases our knowledge about some aspect of it. However, we've seen that any such measurement according to our model will disturb the system to an extent that we cannot monitor; and this decreases our knowledge about some other aspect of the system. For a single measurement this tradeoff is expressed by the precision-disturbance relation (22), or in more detail by the updates (20a, 20b). In this section we explore the compound effect of such tradeoff due to multiple measurements; specifically, a continuous succession of vanishingly-weak measurements. This will allow us, in Section I.5.3, to treat the cases of simultaneous measurement of multiple observables, and of inefficient measurements. The method of analysis we follow is drawn from the field of continuous quantum measurement, which addresses the corresponding problem in that setting. (See for example [16].)

Subdivide a finite interval of time $[0,T]$ into N equal subintervals demarcated by $t_0 = 0 < t_1 < t_2 < \cdots < t_N = T$, with $t_j = j\Delta t$. For each $j \in \{1,\dots,N\}$, select an observable $A_j = A_j(q,p)$ of the system, and prepare for it a measurement $(k_j\Delta t, q_j)$ to be carried out at time t_j. Notice that we've scaled the strength according to the size of the subintervals; smaller Δt means each individual measurement is weaker, but a greater number of them fit into $[0,T]$. We will see that this is the right scaling for the effects to converge when we take the limit of smaller and smaller Δt. (Note that this changes the physical dimensions of k_j; they are now $[k_j] = [A_j]^{-2} \cdot \text{time}^{-1}$.) The resulting tuple of pointer readings $\mathbf{A}^* \triangleq (A_1^*, A_2^*, \dots, A_N^*)$ constitutes the measurement record for the entire succession of measurements. To assimilate the j-th measurement we perform the two operations (20a, 20b), resulting in the update

$$\rho(q,p;t_{j+1}) \propto e^{-4k_j\Delta t\left(A_j^* - A_j(q,p)\right)^2} \int_{-\infty}^{\infty} \frac{d\tau\, e^{-\frac{1}{4(k_j\Delta t)q_j^2}\tau^2}}{\sqrt{4\pi(k_j\Delta t)q_j^2}} \left[\left(\Phi_\tau^{A_j}\right)_* \rho\right](q,p;t_j). \tag{23}$$

As Δt becomes small, the exponential inside the integral vanishes except for small τ. For small τ, the push-forward

$$\left[\left(\Phi_\tau^A\right)_* \rho\right](q_0,p_0) = \rho\left(\Phi_{-\tau}^A(q_0,p_0)\right) = \rho(q(-\tau),p(-\tau)) \tag{24}$$

can be calculated by Taylor-expanding the function $\tau \mapsto \rho(q(-\tau),p(-\tau))$ around $\tau = 0$; using the chain rule to pass all time-derivatives onto q,p, and calculating the latter from

Hamilton's equations with Hamiltonian A. The result is

$$\left(\Phi_\tau^A\right)_* \rho = \rho + \tau\{A,\rho\} + \frac{\tau^2}{2}\{A,\{A,\rho\}\} + \mathcal{O}(\tau^3). \tag{25}$$

Putting this into (23), the integral can then be done order-by-order. The odd-order terms all vanish by symmetry, leaving us with

$$\rho(q,p;t_{j+1}) \propto \exp\left\{-4k_j\Delta t\left(A_j^* - A_j\right)^2\right\}\left(\rho + k_j q_j^2 \Delta t\{A_j,\{A_j,\rho\}\} + \mathcal{O}(\Delta t^2)\right)\Big|_{(q,p;t_j)}. \tag{26}$$

I.5.1 Discarded measurement record

Let's pause to consider the case in which the measurement record $\mathbf{A}^*$ is discarded. In this case we should skip update (20b), which amounts to dropping the exponential factor and the omitted proportionality factor in (26). Taking then the limit $\Delta t \to dt$ describing a continuous succession of vanishingly-weak measurements, we arrive in this case (discarded measurement record) at

$$\frac{\partial\rho}{\partial t} = \underbrace{\{H,\rho\}}_{\substack{\text{internal dynamics}\\ \text{Hamiltonian flow}\\ \text{info preserved}}} + \underbrace{kq^2\{A,\{A,\rho\}\}}_{\substack{\text{observer effect}\\ \text{diffusion along flow } \Phi_\tau^A\\ \text{info of compatible observs. preserved}\\ \text{other info lost}}} \tag{27}$$

where we've introduced the well-known Liouville term $\{H,\rho\}$ accounting for the internal dynamics of the system under H [21], which we had been ignoring until now; and all quantities shown may be explicit functions of time. This is a Liouville-like master equation, with an additional second-order term due to the observer effect of measurement. We can get some sense for the effect of this new term as follows. Let $B(q,p;t)$ denote any function over phase space, possibly explicitly time-dependent. Here and throughout let's use $\langle\cdot\rangle$ to denote the phase-space average:

$$\langle B\rangle \triangleq \int d^n q d^n p\, \rho(q,p;t) B(q,p;t). \tag{28}$$

In Appendix I.A we prove that under master equation (27) any such phase-space average evolves as

$$\frac{d}{dt}\langle B\rangle = \langle\{B,H\}\rangle + \left\langle\frac{\partial B}{\partial t}\right\rangle - kq^2\left\langle\{\{A,\log\rho\}\{A,B\}\}\right\rangle. \tag{29}$$

The first term on the right-hand side of this equation is due to the Liouville term in (27); the second term is due to any explicit time-dependence of B; and the third term is due to the second-order term in (27). As a special application of this equation consider $B = -\log\rho$, in which case the phase-space average is the Gibbs entropy:

$$S(t) \triangleq \langle -\log\rho\rangle. \tag{30}$$

It is not hard to show that the first two terms on the right-hand side of (29) vanish in this case.[8] Thus we find that under dynamics (27),[9]

$$\dot{S} = kq^2 \left\langle \{A, \log\rho\}^2 \right\rangle \geq 0. \tag{31}$$

It is a well-known result that $S(t)$ remains constant ($\dot{S} = 0$) under the Liouville equation $\partial\rho/\partial t = \{H, \rho\}$ (see example in Figure 1b). In breaking with that, we have just found that entropy generally increases over time under (27) on account of the new term. Thus the Liouville term preserves information, while the second-order term causes information loss. Indeed, in accordance with our discussion in Section I.4 concerning the nature of the observer effect, this term describes diffusion along the flow lines generated by the instantaneous observable $A(q, p; t)$ (see example in Figure 1c). This diffusion preserves, instant-to-instant, information pertaining to observables in involution with $A(q, p; t)$, while it erases information pertaining to observables not in involution with it.

We should note that master equation (27) has appeared in the literature before, outside the context of measurement. It appeared in [22], which studied stochastic optimization problems. And a generalization of it appeared in [23], which studied Hamiltonian systems driven by colored noise.[10]

I.5.2 Simulated measurement record

Returning now to (26), suppose instead that the measurement record is not discarded but that we have only yet read up to the $(j-1)$-th entry; i.e. A_1^* through A_{j-1}^* are known while A_j^* onward are not. We would like to simulate ahead of time (say, on a computer) how our state of knowledge will evolve as we continue to read more of the record. However, without the benefit of hindsight the upcoming record entries appear to us as random variables. The language for this kind of simulation is stochastic calculus. (See tutorial on stochastic calculus in [16].) Let us first ask: what should be our probability distribution for the upcoming outcome, A_j^*? Making use of the likelihood function (15), this question can be answered in terms of our current knowledge of the value of A_j:

$$\begin{aligned}\rho(A_j^*; t_j) &= \int dA_j\, \rho(A_j; t_j)\rho(A_j^*|A_j) \\ &\propto \int dA_j\, \rho(A_j; t_j) \exp\left\{-4k_j\Delta t \left(A_j^* - A_j\right)^2\right\}.\end{aligned} \tag{32}$$

As $\Delta t \to dt$, the exponential in this expression becomes very wide and spread out as a function of A_j. The distribution $\rho(A_j; t_j)$ becomes very narrow by comparison, and can be replaced by a Dirac delta, which must be centered at $\langle A_j \rangle$ for the means to match. Using the delta to do the integral over A_j we have, up to a normalization factor,

$$\rho(A_j^*; t_j) \underset{\Delta t \to dt}{\longrightarrow} \exp\left\{-4k_j\Delta t \left(A_j^* - \langle A_j\rangle\right)^2\right\}. \tag{33}$$

[8] Proof: by identity (64) from Appendix I.A, the first term on the right-hand side of (29) can be written as $\int H\{\rho, -\log\rho\}$, which is zero because the bracket vanishes. The second term on the right-hand side of (29) is $\int \rho \frac{\partial}{\partial t}(-\log\rho) = -\int \frac{\partial\rho}{\partial t} = -\frac{d}{dt}\int \rho = -\frac{d}{dt}1 = 0$.

[9] (31) is a special case of a more general result,

$$\frac{d}{dt}\int d^n q d^n p\, f(\rho) = -kq^2 \int d^n q d^n p\, f''(\rho)\{A, \rho\}^2,$$

which holds under dynamics (27). (We omit this result's proof, which involves steps similar to those leading to (31).) Here $f : \mathbb{R} \to \mathbb{R}$ is any smooth function for which the shown integrals converge. It follows that for every such function, f, which is downward-concave, we have an H-theorem, $\frac{d}{dt}\int f(\rho) \geq 0$. (31) is the case $f(\rho) = -\rho\log\rho$.

[10] I thank an anonymous referee for pointing out these connections.

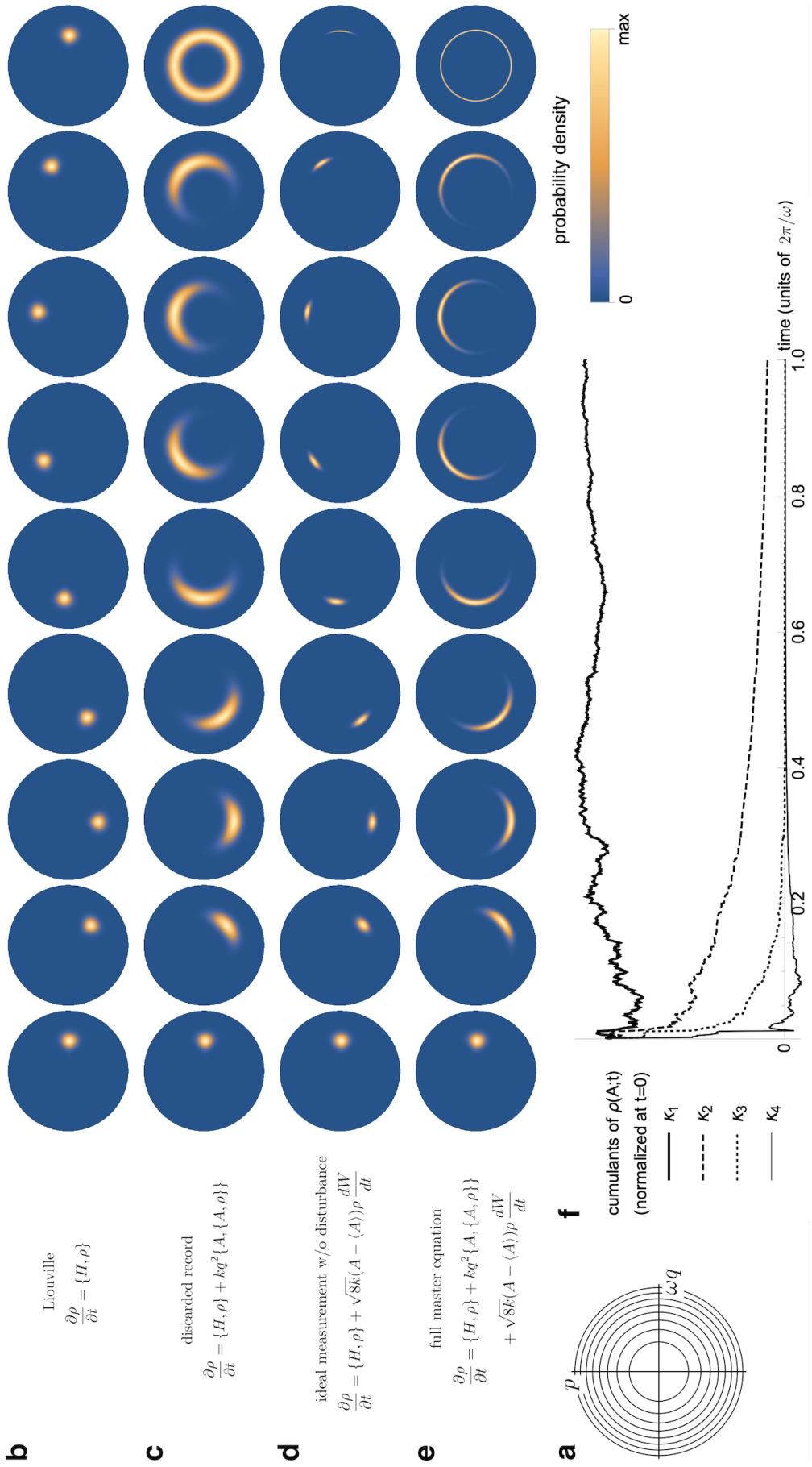


Figure 1: (Rotated. Caption next page.)

Figure 1: (Previous page.) **Master equation dynamics in various measurement regimes.** Evolution of the state of knowledge $\rho(q,p;t)$ of a rational agent under master equation (39) is illustrated in a simple example: the system under measurement is a 1D simple harmonic oscillator (sho); the measurement is characterized by constant k, q and fixed A; the measured observable is the energy $A = H \triangleq \frac{1}{2}(\omega^2 q^2 + p^2)$; and the initial distribution over phase space is unimodal. Although not proven here, three timescales are involved: that of internal dynamics, $\tau_{\text{dyn}} \sim 1/\omega$; that of diffusion due to observer effect, $\tau_{\text{dif}} \sim 1/kq^2\omega^2$; and that of collapse due to Bayesian update on the measurement record, $\tau_{\text{col}} \sim 1/k\Delta E^2$, where ΔE is the target certainty on H (i.e. τ_{col} is the characteristic timescale for the variance of $\rho(H;t)$ to fall below ΔE^2). **(a)** Phase portrait showing level sets of the sho Hamiltonian. **(b–e)** Snapshots of $\rho(q,p;t)$ at successive times, indicated at top in units of the sho period, for four different measurement regimes (rows). The simplified master equation in each regime is indicated at left. For ease of visualization the color scheme (bottom right) is normalized anew for each plot. **(b)** Regime $\tau_{\text{dyn}} \ll \tau_{\text{dif}}, \tau_{\text{col}}$; describes an isolated system; (39) reduces to the Liouville equation $\partial\rho/\partial t = \{H, \rho\}$. **(c)** Regime $\tau_{\text{dyn}} \sim \tau_{\text{dif}} \ll \tau_{\text{col}}$; describes case of discarded measurement record; (39) reduces to (27). Notice entropy increase, in accordance with (31), due to diffusion along the flow generated by A. **(d)** Regime $\tau_{\text{dyn}} \sim \tau_{\text{col}} \ll \tau_{\text{dif}}$; describes an approximation to ideal classical measurement with minimal disturbance. Notice the trend of decreasing entropy, in accordance with (42), due to collapse towards the measurement outcome. **(e)** Regime $\tau_{\text{dyn}} \sim \tau_{\text{col}} \sim \tau_{\text{dif}}$; describes the three processes (dynamics, diffusion and collapse) happening together. Notice the tradeoff between information about A and information about the conjugate quantity (sho phase). **(f)** Evolution of the first four cumulants of $\rho(A;t)$ in regime **d** (equivalently regime **e**). For ease of visualization each cumulant is rescaled to 1 at $t = 0$. Note qualitative agreement with (44).

By a simple change of variables we introduce ΔW_j, our probability distribution of which is a zero-mean Gaussian with variance Δt, and in terms of which

$$A_j^* = \langle A_j \rangle + \frac{1}{\sqrt{8k_j}} \frac{\Delta W_j}{\Delta t}. \tag{34}$$

The value of expressing (33) this way is two-fold. From a simulation standpoint, we can use a random number generator to sample ΔW_j from its Gaussian distribution, and (34) then tells us how to convert this into a sample of A_j^*. And from an analysis standpoint, this expression enables a very convenient form of calculation: in the limit $\Delta t \to dt$ we write

$$A^* = \langle A \rangle + \frac{1}{\sqrt{8k}} \frac{dW}{dt}, \tag{35}$$

where $W(t) \triangleq \int_0^t dW$ is a standard Wiener process, with dW obeying the basic rule of Itô calculus $dW^2 = dt$. Notice that ΔW_j is statistically-independent from all quantities appearing up to time $t = t_j$. Using $\langle\!\langle \cdot \rangle\!\rangle$ to denote averaging over the Wiener process, we have in particular, for any function $f(\rho, A)$ of the present ρ and A:

$$\langle\!\langle f(\rho, A) dW \rangle\!\rangle = f(\rho, A) \langle\!\langle dW \rangle\!\rangle = 0. \tag{36}$$

Taking stock: given $\rho(q,p;t_j)$ for a given time t_j we can use it to calculate $\langle A_j \rangle$ (as in (28)), and combine this with the output of a random number generator as in (34) to simulate the upcoming entry of the measurement record A_j^*. We can then use (26) to calculate what our updated state of knowledge $\rho(q,p;t_{j+1})$ would be upon reading that entry, and iterate the process. Analytically we proceed as follows. Substitute (34) into (26); expand the square

in the exponent, discarding the overall factor $\exp\{-\Delta W_j^2/2\Delta t\}$ which is independent of (q,p); and Taylor-expand the exponential, keeping in mind that powers of ΔW_j count for "half an order", to obtain

$$\rho(q,p;t_{j+1}) \propto \Big(1 - 4k_j\Delta t(A_j - \langle A_j\rangle)^2 + \sqrt{8k_j}\Delta W_j(A_j - \langle A_j\rangle) + 4k_j\Delta W_j^2(A_j - \langle A_j\rangle)^2 + \mathcal{O}(\Delta t\,\Delta W_j)\Big)\Big(\rho + k_j q_j^2\Delta t\{A_j,\{A_j,\rho\}\} + \mathcal{O}(\Delta t^2)\Big)\Big|_{(q,p;t_j)} . \tag{37}$$

In the limit of continuous measurement $\Delta t \to dt, \Delta W_j \to dW, \Delta W_j^2 \to dt$ this reduces to

$$\rho(q,p;t+dt) \propto \rho + kq^2\{A,\{A,\rho\}\}dt + \sqrt{8k}(A - \langle A\rangle)\rho\, dW\Big|_{(q,p;t)} , \tag{38}$$

where again all quantities shown may be explicit functions of time. One can check that the right-hand side is already normalized, so the omitted factor of proportionality is 1. We arrive in this case (simulated measurement record) at

$$\frac{\partial\rho}{\partial t} = \underbrace{\{H,\rho\}}_{\substack{\text{internal dynamics}\\ \text{Hamiltonian flow}\\ \text{info preserved}}} + \underbrace{kq^2\{A,\{A,\rho\}\}}_{\substack{\text{observer effect}\\ \text{diffusion along flow } \Phi_\tau^A\\ \text{info of compatible observs. preserved}\\ \text{other info lost}}} + \underbrace{\sqrt{8k}(A - \langle A\rangle)\rho\,\frac{dW}{dt}}_{\substack{\text{Bayesian update}\\ \text{collapse towards measurement outcome}\\ \text{non-linear \& non-local}\\ \langle\!\langle\Delta\text{info}\rangle\!\rangle\geq 0}} , \tag{39}$$

where again we've re-introduced the Liouville term $\{H,\rho\}$ accounting for the internal dynamics of the system. Compared to (27) we now have a new stochastic term appearing, which is due to assimilation of the measurement record via Bayesian update. It is interesting to note that this term is both non-linear and non-local in ρ, since $\langle A\rangle$ depends on the value of ρ everywhere on phase space. To get some sense for the effect of this new term, in Appendix I.B we prove that under master equation (39) the Gibbs entropy (30) evolves as

$$\dot S = \underbrace{kq^2\left\langle\{A,\log\rho\}^2\right\rangle}_{\substack{\text{observer effect}\\ \Delta\text{entropy}\geq 0}} \underbrace{-4k\sigma_A^2 - \sqrt{8k}\left\langle(A - \langle A\rangle)\log\rho\right\rangle\frac{dW}{dt}}_{\substack{\text{Bayesian update}\\ \text{can be positive or negative}}}, \tag{40}$$

where

$$\sigma_A^2 = \sigma_A(t)^2 \triangleq \langle(A - \langle A\rangle)^2\rangle \tag{41}$$

is the variance in our knowledge of $A(q,p;t)$ at time t. The first term on the r.h.s. of (40) is familiar from (31); it describes increasing entropy due to the observer effect of measurement. The remaining two terms are due to the stochastic term in (39); these two together may be positive for particular measurement outcomes, but they are non-positive on average, as can be seen by invoking (36):

$$\langle\!\langle\dot S\rangle\!\rangle = \underbrace{kq^2\left\langle\{A,\log\rho\}^2\right\rangle}_{\substack{\text{observer effect}\\ \Delta\text{entropy}\geq 0}} \underbrace{-\ 4k\sigma_A^2}_{\substack{\text{Bayesian update}\\ \langle\!\langle\Delta\text{entropy}\rangle\!\rangle\leq 0}} . \tag{42}$$

Thus the stochastic term in (39) leads, on average, to increasing information (see example in Figure 1d,e). To gain further insight into the effects of this term, suppose the measured

observable is fixed $A = A(q,p)$, and consider our PDF over this observable, $\rho(A;t)$, which is just the marginal

$$\rho(A';t) \triangleq \int d^n q d^n p\, \delta(A(q,p) - A')\rho(q,p;t). \tag{43}$$

Let κ_i denote the i-th cumulant of this distribution. In Appendix I.C we prove the following hierarchy of equations describing the contribution of the stochastic term in (39) to the evolution of these cumulants:

$$d\kappa_1 = \sqrt{8k}\,\kappa_2\, dW, \tag{44a}$$
$$d\kappa_2 = \sqrt{8k}\,\kappa_3\, dW - 4k(2\kappa_2^2)dt, \tag{44b}$$
$$d\kappa_3 = \sqrt{8k}\,\kappa_4\, dW - 4k(6\kappa_2\kappa_3)dt, \tag{44c}$$
$$d\kappa_4 = \sqrt{8k}\,\kappa_5\, dW - 4k(8\kappa_2\kappa_4 + 6\kappa_3^2)dt, \tag{44d}$$
$$\dots$$

Notice in particular the trends $\langle\!\langle \dot\kappa_1 \rangle\!\rangle = 0$, $\langle\!\langle \dot\kappa_2 \rangle\!\rangle \sim -\langle\!\langle \kappa_2 \rangle\!\rangle^2$, $\langle\!\langle \dot\kappa_3 \rangle\!\rangle \sim -\langle\!\langle \kappa_3 \rangle\!\rangle$, $\langle\!\langle \dot\kappa_4 \rangle\!\rangle \sim -\langle\!\langle \kappa_4 \rangle\!\rangle$, ... These trends tell us that (supposing A is not explicitly time-dependent and the Liouville term does not intervene too strongly) the stochastic term in (39) causes all cumulants of $\rho(A;t)$ higher than second to vanish exponentially fast, leaving $\rho(A;t)$ a Gaussian; it then causes the variance to vanish like $\sim 1/t$, while the mean jiggles around in a random walk of zero drift and volatility decaying with the variance. In the limit in which the measurement process is complete, $\rho(A;t)$ converges to a delta distribution centered at the simulation's putative true value of A. (See example in Figure 1d–f.)

I.5.3 Simultaneous and inefficient measurements

Simultaneous weak measurement of multiple observables $A_1(q,p), \dots, A_s(q,p)$, whether these are in involution or not, can be handled by letting $A(q,p;t)$ in (39) switch between these observables on a fast time scale. Inefficient measurements can be handled in this way too, by sporadically (on the fast time scale) discarding some of the outcomes some of the time, thus reducing (39) to (27) at those times. By averaging the resulting dynamics over the fast time scale we're left with

$$\frac{\partial\rho}{\partial t} = \{H,\rho\} + \sum_{j=1}^{s} k_j q_j^2 \{A_j, \{A_j,\rho\}\} + \sum_{j=1}^{s} \sqrt{8\nu_j k_j}(A_j - \langle A_j\rangle)\rho\,\frac{dW_j}{dt}, \tag{45}$$

where (k_j, q_j, ν_j) describes the measurement setup for the j-th observable, and $W_j(t) \triangleq \int_0^t dW_j$ are independent Wiener processes for $j \neq j'$. Here $\nu_j \in [0,1]$ is the *efficiency* of the j-th measurement. A perfectly efficient measurement has $\nu = 1$ (as in (39)), while a perfectly inefficient measurement has $\nu = 0$ and corresponds to discarding the outcome (as in (27)).

The analogues of (40) and (42) for the above equation are

$$\dot S = \sum_{j=1}^{s} k_j q_j^2 \left\langle \{A_j, \log\rho\}^2 \right\rangle - \sum_{j=1}^{s} \left(4\nu_j k_j \sigma_{A_j}^2 + \sqrt{8\nu_j k_j}\left\langle (A_j - \langle A_j\rangle)\log\rho \right\rangle \frac{dW_j}{dt} \right), \tag{46}$$

and

$$\langle\!\langle \dot S \rangle\!\rangle = \sum_{j=1}^{s} k_j q_j^2 \left\langle \{A_j, \log\rho\}^2 \right\rangle - \sum_{j=1}^{s} 4\nu_j k_j \sigma_{A_j}^2. \tag{47}$$

If all the outcomes are discarded ($\nu_j = 0$ for all j) we're left with

$$\frac{\partial\rho}{\partial t} = \{H,\rho\} + \sum_{j=1}^{s} k_j q_j^2 \{A_j, \{A_j,\rho\}\}, \tag{48}$$

which is linear, local, and deterministic; and

$$\dot{S} = \sum_{j=1}^{s} k_j q_j^2 \left\langle \{A_j, \log \rho\}^2 \right\rangle \geq 0. \tag{49}$$

I.6 Discussion

I.6.1 Comparing the quantum and classical uncertainty relations

How does our classical precision-disturbance relation (22) compare to the Heisenberg uncertainty principle of quantum mechanics? The latter can be stated in a few different forms. We will consider the Kennard-Weyl-Robertson form in section I.6.2, where we discuss the epistemology of classical Hamiltonian ontology. Here we consider the "joint measurement form" [24–27], pertaining to simultaneous measurement of two observables, A and B. When A, B are conjugate to each other this reads:

$$\epsilon_A \epsilon_B \geq \frac{\hbar}{2}, \tag{50}$$

where ϵ_A and ϵ_B denote the imprecisions in the measurement of A and B, respectively;[11] and $\hbar$ is the reduced Planck constant.

One superficial difference between (22) and (50) is that one is an equality while the other an *in*equality. However, this difference is illusory. The product on the left-hand side of (22) can easily be made larger than the right-hand side, so that for a more general class of measurement models we have

$$\epsilon_A \eta_A \geq \frac{q}{2}. \tag{51}$$

Indeed, we have defined inefficient weak measurements in Section I.5.3, as those with $\nu < 1$ in (45). Such measurements will fail to saturate (51). The extreme case of this is when the measurement outcome is discarded ($\epsilon_A \to \infty; \eta_A$ unchanged). Another way to modify our measurement model that fails to saturate (51) is if the apparatus' pointer fails to be in involution with $H^{\text{own}}_{\text{app}}$, so that some amount of "deterioration" of the measurement record can happen between the time of the system-apparatus interaction and whenever the record is read. In the opposite direction, one might ask: could not the bound (51) be exceeded, say, by using a coupling and pointer, (Q, P), that are correlated in the apparatus ready state? To achieve the latter, one would need (Q, P) to *not* diagonalize the quadratic form (9); namely, instead of choosing $(Q, P) = (z_{i^*}, w_{i^*})$, one would choose (Q, P) related to (z_{i^*}, w_{i^*}) by some linear canonical transformation. In fact, although not proven here, I find that this approach leads to the same precision-disturbance relation (22); the only difference (aside from the Bayesian analysis becoming more involved) is that a systematic component is added to the observer effect. This component can be corrected given the measurement outcome A^*; so it doesn't count towards the disturbance η_A.[12] In summary, these remarks suggest that, while it is easy to do worse than (22), it may not be possible to do better; i.e. they suggest inequality (51) to be the general principle.

A second difference, which remains between (50) and (51), is that one involves the product of two imprecisions, while the other the product of an imprecision with a disturbance magnitude. This difference can be bridged as well. Recall that the disturbance in question amounts to flowing along Φ^A_τ for an unknown "time" τ whose uncertainty is η_A. Under

[11] To ease comparison between the quantum and classical cases, it's convenient to speak of quantum mechanics from a realist/hidden-variable interpretation, in which measurement outcomes are outcomes *about* an underlying unknown state. We won't concern ourselves here with the ongoing debate about the plausibility of such an interpretation.

[12] Recall: η_A is defined as our *uncertainty* in τ (τ being the "flow time" for which the measurement caused the system to move along Φ^A_τ).

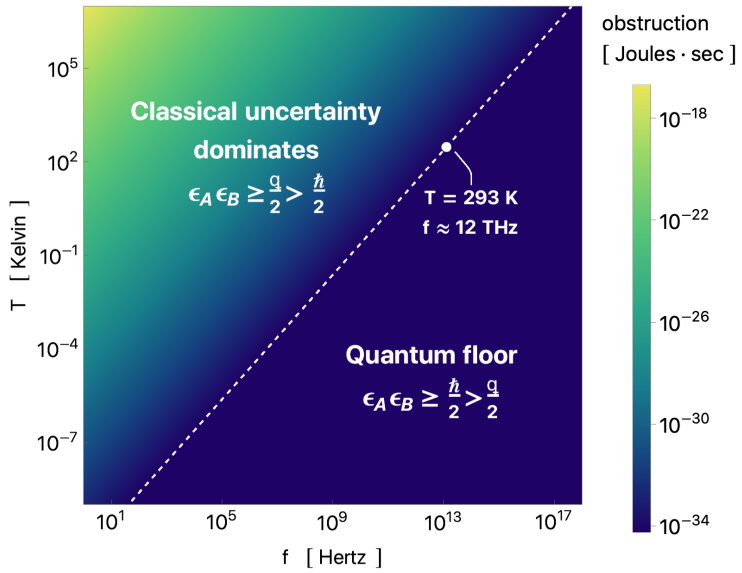


Figure 2: **Coexistence of quantum and classical uncertainty relations in the real world.** One may expect the quantum and classical uncertainty relations to coexist in reality based on the observation that a Hamiltonian world effectively emerges from quantum mechanics at macroscopic scales. The quantum relation dominates at low apparatus temperature and/or tightly-trapped pointer in the apparatus ready state; when $q < \hbar$ (below the dashed diagonal). The classical relation dominates in the other direction. Here $f = \Omega/2\pi$ and $T = 1/k_B\beta$. Notice that for the range of (f, T) shown, the obstruction is never larger than $\sim 10^{-17}$J·s.

this flow the "rate" of change of any observable B is as given by (2): $\frac{d}{d\tau}B = \{B, A\}$. In particular, if B is the conjugate to A, so that $\{B, A\} = 1$, then B increases monotonically at the steady rate of 1; and the net effect of the flow on B is simply to displace its value by τ. (This final step can fail if B has a discontinuity somewhere; so it is important that B be a bona fide observable, not just a local quantity such as the phase ϕ of an oscillator.) So the uncertainty in the "flow time", η_A, translates directly into a disturbance in the value of the conjugate observable, B. This places a lower bound on the imprecision, ϵ_B, with which any subsequent measurement can hope to determine the original value of B: $\epsilon_B \geq \eta_A$, with equality holding only if the measurement of B is done at full strength ($k \to \infty$). Thus we have

$$\epsilon_A \epsilon_B \geq \frac{q}{2}, \tag{52}$$

and the parallel with (50) becomes apparent. Historically it seems that Heisenberg's own interpretation of the uncertainty principle was as a precision-disturbance relation [2], not very different in spirit from (51). And in recent years work in quantum mechanics has payed considerable attention to precision-disturbance relations, yielding formulas similar to (51) (with $\hbar$ in place of q) [1, 28–31].

The real world is no doubt quantum mechanical, and so the Heisenberg uncertainty principle is fundamental. But as we know, as one "zooms out" to larger scales somehow an approximately Hamiltonian world effectively emerges (Bohr's correspondence principle and the quantum-to-classical transition). Hand in hand with the emergence of this effective Hamiltonian world I expect our classical uncertainty relation to gain traction. Figure 2 illustrates how the classical and quantum relations then must coexist. For a tight enough trap and/or cold enough apparatus (below the dashed diagonal), the obstruction in (52) is brought below $\hbar/2$ and becomes unreachable; the quantum obstruction acts like rock bottom. For less tight traps and/or warmer apparatuses the obstruction in (52) rises

above $\hbar/2$ and begins to dominate. Taken together, one may expect to have in the real world an obstruction that interpolates between these two; something along the lines of

$$\epsilon_A \epsilon_B \geq \frac{\hbar + q}{2} \quad \text{or perhaps} \quad \frac{\hbar/2}{1 - e^{-\hbar/q}}; \tag{53}$$

it will take a detailed quantum calculation to work out the precise formula (see Section I.6.4 for a germ of how this might be done). To gain some perspective for the scales involved, note from Figure 2 that, at room temperature, trap frequencies any lower than about 12THz (corresponding to light wave-lengths $\gtrsim 25\mu$m) are already enough to put us in the classical regime. At the same time, even for the highest temperatures and lowest frequencies shown in the top-left of Figure 2, the classical obstruction hardly becomes larger than $\sim 10^{-17}$J·s; an extremely small quantity by macroscopic standards. And yet, even in more moderate regimes towards the center of Figure 2, the classical obstruction may be relevant in the contexts of precision measurement, nanoengineering and molecular machines.

I.6.2 On the epistemology of classical Hamiltonian ontology

Consider the Kennard-Weyl-Robertson (KWR) form of the Heisenberg uncertainty principle of quantum mechanics [32]. For a pair of conjugate observables, A and B, it reads:

$$\sigma_A \sigma_B \geq \frac{\hbar}{2}, \tag{54}$$

where σ_A and σ_B denote the standard deviations at a given time in our knowledge of A and B, respectively.[13] This form of the uncertainty principle speaks directly to the limits of what can be known about the state of a quantum system; that is, to the epistemology of quantum ontology. In this section we ask whether the present developments allow us to establish an analogous result about the epistemology of classical Hamiltonian ontology.

Notice, first of all, the sense in which we must understand such a question. Unlike in the quantum formalism, there is nothing in our classical formalism that rules out the possibility of *starting* with perfect information about conjugate observables:

$$\rho(A', B'; t) = \delta(A(t) - A')\delta(B(t) - B'), \tag{55}$$

where

$$\rho(A', B'; t) \triangleq \int d^n q d^n p \, \delta(A(q,p) - A')\delta(B(q,p) - B')\rho(q,p;t). \tag{56}$$

Rather, the question is whether it is at all possible to *arrive* at such a state of perfect information from a state of less information. In particular: suppose we were handed a Hamiltonian system about whose state we knew nothing at all, so that ρ were initially uniform on phase space. Does there exist a sequence of measurements on the system that would take ρ into the perfect-information state (55)?

Consider the direct approach of performing simultaneous measurement of A and B, with respective measurement settings (k_A, q_A) and (k_B, q_B), and perfect efficiencies $\nu_A = \nu_B = 1$. The evolution of ρ is as given by master equation (45). Suppose that the measurements are strong enough that they come to completion on a much faster timescale than that of the system's dynamics, so that the Liouville term in (45) can be neglected. It is a bit tricky, because one must be mindful of the rules of Itô calculus, but one can check that, starting from an uncorrelated Gaussian distribution in A and B (of which the uniform distribution is the special case of infinite variances), the general solution to (45) is

$$\rho(A, B; t) = \frac{1}{2\pi\sigma_A\sigma_B} \exp\left\{-\frac{(A-\mu_A)^2}{2\sigma_A^2} - \frac{(B-\mu_B)^2}{2\sigma_B^2}\right\}, \tag{57}$$

[13] cf. footnote 11.

where the means μ_A, μ_B are stochastic functions of time evolving as

$$d\mu_A = \sqrt{8k_A}\sigma_A(t)^2 dW_A, \tag{58a}$$

$$d\mu_B = \sqrt{8k_B}\sigma_B(t)^2 dW_B; \tag{58b}$$

while the variances σ_A^2, σ_B^2 are the deterministic functions of time

$$\sigma_A(t)^2 = \frac{q_B}{2}\sqrt{\frac{k_B}{k_A}}\left[\coth\left\{4q_B\sqrt{k_A k_B}(t-t_A)\right\}\right]^{l_A}, \tag{59a}$$

$$\sigma_B(t)^2 = \frac{q_A}{2}\sqrt{\frac{k_A}{k_B}}\left[\coth\left\{4q_A\sqrt{k_A k_B}(t-t_B)\right\}\right]^{l_B}, \tag{59b}$$

where $t_A, t_B < t$ are constants of integration, as are $l_A, l_B \in \{+1, -1\}$. We see that, under simultaneous measurement of conjugate observables, an initially-Gaussian-uncorrelated PDF remains so for all time. Also, much like we saw in (44), the mean of the distribution executes a random walk (this time in two dimensions) of volatilities proportional to the variances. However, unlike in (44), now the variances converge to non-zero values as the measurements run to completion ($t \to \infty$):

$$\sigma_A^2 \to \frac{q_B}{2}\sqrt{\frac{k_B}{k_A}}, \qquad \sigma_B^2 \to \frac{q_A}{2}\sqrt{\frac{k_A}{k_B}}. \tag{60}$$

This comes about because the measurement of A causes collapse "along the A-direction" (along the integral curves of Φ_τ^B) and diffusion "perpendicular to the A-direction" (along the integral curves of Φ_τ^A); while the simultaneous measurement of B causes the converse; and at completion of the measurement the effects precisely cancel out. Notice that (60) gives

$$\sigma_A\sigma_B \to \frac{\sqrt{q_A q_B}}{2}, \tag{61}$$

which begins to resemble (54). Is it the case that the product $\sigma_A(t)\sigma_B(t)$ remains above this limit at all times? That depends on the exponents l_A, l_B. The case $l_A = +1$ gives $\sigma_A(t_A^+) \to \infty$; it describes complete ignorance about A at some past time t_A. On the other hand, the case $l_A = -1$ gives $\sigma_A(t_A^+) = 0$; it describes perfect information about A at the past time t_A. Likewise for l_B.[14] Since we are interested in beginning from a state of ignorance, the relevant solution for us has $l_A = l_B = +1$. It then follows from (59) that the inequality

$$\sigma_A\sigma_B \geq \frac{\sqrt{q_A q_B}}{2} \tag{62}$$

holds for all times. If both measurement apparatuses are of the same (inverse) quality, q, this further reduces to

$$\sigma_A\sigma_B \geq \frac{q}{2}. \tag{63}$$

We have derived this uncertainty-uncertainty relation by considering simultaneous measurement of the pair of conjugate observables A, B. Could this be a general epistemic obstruction, or is there some different sequence of measurements that fares better? We leave the question open for future investigation.

[14] It is noteworthy that when σ_A and σ_B are smaller than their terminal values (60) (i.e. for $l_A = l_B = -1$) one actually *loses* information by measuring! (Because the induced disturbances win over the information gains.)

I.6.3 Reading the measurement record; is it turtles all the way down?

We return to a complication that was tacitly overlooked in Section I.3. In our measurement model, after the apparatus had interacted with the system—let's call that step "pre-measurement"—we stipulated that the pointer on the apparatus should be read, which would yield some definite value P^*. But what could it mean to "read P" if not to measure this observable of the apparatus? This seems to lead us down an infinite regression in which the system is pre-measured by an apparatus, which must then be pre-measured, presumably by another apparatus, which must then be.... The passage from "systems interacting" to "agent being informed" never quite taking place. In a sense this predicament is similar to the quantum measurement problem, particularly as articulated by the Wigner's friend thought experiment [33]. In both settings, the paradoxical step is the passage from what seems to be best regarded as an ontic-level description (Hamiltonian or unitary dynamics) to what seems to be best regarded as an epistemic-level operation (Bayesian updating or collapse of the wave function). At the epistemic level we speak freely of agents, observers, measurements, observations, reading the measurement record, information, probability, Bayesian updating, collapse of the wave function. But at the ontic level all of these are complicated phenomena, resisting precise characterization. Until such characterizations are available we are stuck with the *shifty split*—to use a term coined by Bell [34].

In our model, the shifty split was introduced at one degree of separation from the system under study: we described the system-apparatus interaction at the ontic level; then we described the apparatus-agent interaction at the epistemic level. Such a once-removed approach can be very useful, as demonstrated by the example of the theory of general quantum measurement [35]. However, there is a pair of consistency tests that should be checked of such a theory. Notice that the once-removed theory contains the twice-removed and higher theories. To see this, simply let what we have been calling the object-system instead be used as an apparatus of some kind to (pre-)measure another system. (Now we have an object-system which is pre-measured by an apparatus, which is pre-measured by a second apparatus, whose pointer is "read" by an agent.) This maneuver uses only the rules of the ontology, yet succeeds in shifting away the split by one degree. In light of this feature of the theory, for a first test (T1) we ask: is there any such way to shift away the split that increases the efficiency of our measurement (i.e. decreases the obstruction in (51))? If the answer is yes then test (T1) is failed; it is a sign that the way we are bridging the shifty split does not fully exploit the possibilities allowed by the ontology, and we should strengthen it. (This strengthening is necessary. So long as test (T1) is not passed, bounds such as (51) derived from once-removed measurement schemes cannot be taken seriously, since they can be circumvented by better use of the operations allowed by the ontology.) On the other hand, for a second test (T2) we ask: is it the case that when we shift away the split, no matter how we do it, we find that it always decreases the efficiency of our measurement (i.e. increases the obstruction in (51))? If the answer is yes then test (T2) is failed; it is a sign that the way we are bridging the shifty split requires an operation that is not allowed by the ontology, and must be revised.

How does our model fare on these tests? Consider test (T2) first. We have bridged the shifty split by stipulating: (i) "read" the pointer on the apparatus, yielding some definite value P^*. (This is after the apparatus has pre-measured the object-system.) Here's a way to shift away the split without reducing efficiency. Instead of (i) do: (ii) use a second apparatus, operated according to our same model with parameters $(k^{(2)}, q^{(2)})$, to pre-measure at full strength $(k^{(2)} \to \infty)$ the value of P, recording it on its own pointer $P^{(2)}$ which we then "read", yielding some definite value $P^{(2)*}$. To see that procedure (ii) is just as efficient as (i) notice, first, that neither of the two involve further disturbance of the object-system, so η_A is the same for both. Second, since the final measurement in procedure (ii) is at full strength, we have $\epsilon_P^{(2)} \propto 1/\sqrt{k^{(2)}} \to 0$, so this measurement reveals the exact

value of P; $P^{(2)*} = P^*$. Thus procedure (ii) leads to the same likelihood function (15) as (i), and hence to the same ϵ_A. This gives us proof-by-example that it is possible to shift away the split without reducing our model's efficiency, so test (T2) is passed. It is harder to prove that test (T1) is passed since this requires proving that efficiency remains unimproved *for all* ways of shifting away the split. I don't know how to do that, but I conjecture that our model passes this test too. In support of this conjecture, note one way in which efficiency might have been increased by shifting away the split, but isn't. Turn again to procedure (ii) just discussed. Imagine that, after having measured the pointer of the first apparatus (P) as described, it were possible to do another measurement on this apparatus to determine Q. If we could gain even a little information about the value that Q had at the time of interaction with the object-system (t_0), we could combine it with what we know of $Q(t_0)$ from the thermal distribution (11) to reduce our uncertainty σ_Q, and hence reduce the disturbance $\eta_A = \alpha\sigma_Q$. If this were possible our model would fail test (T2). That it is impossible follows from the fact that the measurement of P in procedure (ii) had to be done at full strength ($k^{(2)} \to \infty$), which leads to an infinite disturbance of the first apparatus ($\eta_P^{(2)} \propto \sqrt{k^{(2)}} \to \infty$); and this infinitely disturbs the value of Q (since $\{Q, P\} = 1$). So we see that in the course of carrying out procedure (ii) all information about $Q(t_0)$ is lost beyond recovery.

I.6.4 Future directions

In closing we look to some of the many questions and possibilities ahead. We have argued that our model of measurement is maximally efficient; i.e. that it is impossible to do better than (51) in the way of measuring without disturbing. It would be very desirable to have a proof of this claim at the level of rigor of mathematical physics. Complementing this, it would be desirable to see experimental tests of (51, 52, 63) and of the bigger picture outlined in Figure 2. Moving forward it will be worth honing our intuition about the range of possible dynamics of a Hamiltonian system under measurement (or more accurately, of the epistemic state of a rational agent about a Hamiltonian system under measurement). For this it would be good to see numerical studies of equations (27, 39) in more interesting scenarios than the one-dimensional simple harmonic oscillator explored in Figure 1. For this task it might be useful to carry out in the classical setting the steps, discovered already in the quantum setting, to transform a non-linear stochastic master equation like (39) into an equivalent linear equation [16, 36]. Another calculation that I would like to see is the derivation of the precise version of (53), which I suspect can be obtained by reproducing our measurement model from Section I.3 in the quantum setting. That is, essentially, the von Neumann measurement scheme but taking into account the temperature of the apparatus.

In connection with the quantum measurement problem and the interpretation of quantum mechanics, there is a program dating back to Einstein [37, 38] of attempting to identify and unmix a possible epistemic component of quantum theory from its ontic content. In recent times this program has made promising progress at the hands of Caves, Fuchs, and others [38–40]. In particular Spekkens [41, 42], and Bartlett, Rudolph, and Spekkens [43], have illustrated how an uncircumventable epistemic limitation in an otherwise classical world, much like what is suggested by our discussion in Section I.6.2, can lead to several of the phenomena usually regarded as characteristic of quantum mechanics. It will be interesting to see what these two programs can contribute to each other.

Finally I would like to venture the following speculative suggestions. (i) As we know from general relativity, gravity couples directly to energy. Perhaps a system subject to a strong external gravitational field can, in certain cases, be reasonably modeled by (27) with $A = H$. If so, could this tell us something interesting about the entropy of a system falling onto the event horizon of a black hole? Could this be a useful tool for studying black hole

thermodynamics? (A quantum version of this master equation (c.f. earlier comments in connection to (53)) might be an even better tool.) (ii) To the best of my knowledge, theoretical computer science grounds its notions of computability and complexity in concrete (if highly abstracted and idealized) physical models. Does the existence of an obstruction to ideal measurement without disturbance in Hamiltonian mechanics have a bearing on those notions of computer science grounded in the world of classical physics?[15] (iii) Hamilton's equations and their underlying geometro-algebraic structure are not unique to physics; they emerge wherever the equations of a theory can be gotten out of a variational principle [45]. Indeed, in classical physics they emerge in just this way from Hamilton's principle of stationary action. In particular, optimal control theory uses essentially the same equations under the name of Pontryagin's minimum principle [46]. Could the present results have consequences for aspects of optimal control under partial information and, by extension, for artificial intelligence? (For more on this see Part III.) At the least, these musings illustrate the breadth of potential implications of our subject.

I.A Derivation of equation (29)

Our objective is to derive equation (29). For brevity of notation we will omit the integration measure $d^n q d^n p$ in integrals over phase space. We will make use of the identity

$$\int A\{B,C\} = \int B\{C,A\} = \int C\{A,B\}, \tag{64}$$

which is valid for any smooth functions $A(q,p;t), B(q,p;t), C(q,p;t)$ as long as their product decays to zero as $\|(q,p)\| \to \infty$, so that boundary terms from integration by parts can be discarded. This identity is readily verified:

$$\begin{aligned}
\int A\{B,C\} &= \int A \sum_{i=1}^{n} \left(\frac{\partial B}{\partial q_i}\frac{\partial C}{\partial p_i} - \frac{\partial B}{\partial p_i}\frac{\partial C}{\partial q_i} \right) \\
&= \int C \sum_i \left(-\frac{\partial}{\partial p_i}\left(A\frac{\partial B}{\partial q_i} \right) + \frac{\partial}{\partial q_i}\left(A\frac{\partial B}{\partial p_i} \right) \right) \\
&= \int C \sum_i \left(-\frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i} + \frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i} \right) \\
&= \int C\{A,B\}.
\end{aligned} \tag{65}$$

In our applications of the identity one of the factors will always be homogeneous in ρ, which it is safe to assume decays fast enough for the identity to hold (e.g. for each t, $\rho(q,p;t)$ can be assumed to have compact support over phase space without any loss of physical generality.)

Now, the phase-space average of B is $\langle B \rangle = \int \rho B$, and the time derivative of this is

$$\frac{d}{dt}\langle B \rangle = \int \left(B\frac{\partial \rho}{\partial t} + \rho \frac{\partial B}{\partial t} \right) = \int B\frac{\partial \rho}{\partial t} + \left\langle \frac{\partial B}{\partial t} \right\rangle. \tag{66}$$

Working with the first term on the r.h.s. here, we substitute into it from (27):

$$\int B\frac{\partial \rho}{\partial t} = \int B\left(\{H,\rho\} + kq^2\{A,\{A,\rho\}\} \right). \tag{67}$$

Using identity (64), the first term on the r.h.s. here can be written as $\int \rho\{B,H\} = \langle\{B,H\}\rangle$. Turning to the remaining term on the r.h.s. of (67), we let $C \triangleq \{A,\rho\}$ and again use

[15] Landauer's work [14] establishing the thermodynamic irreversibility of certain computing processes has certainly had such an impact; launching the field of reversible computing [44].

identity (64), so that the integral in this term can be written as

$$\begin{aligned}\int B\{A,C\} &= \int C\{B,A\} = -\int \{A,\rho\}\{A,B\} = -\int \rho\{A,\log\rho\}\{A,B\} \\ &= -\langle\{A,\log\rho\}\{A,B\}\rangle\,. \end{aligned} \tag{68}$$

All together we have

$$\frac{d}{dt}\langle B\rangle = \langle\{B,H\}\rangle + \left\langle\frac{\partial B}{\partial t}\right\rangle - kq^2\,\langle\{A,\log\rho\}\{A,B\}\rangle\,, \tag{69}$$

which is (29), as desired.

I.B Derivation of equation (40)

Our objective is to derive equation (40). For brevity of notation we will omit the integration measure $d^n q d^n p$ in integrals over phase space. Expanding the differential of S (from (30)) to second order in $d\rho$:

$$\begin{aligned} dS &= -\int d(\rho\log\rho) \\ &= -\int \big((\rho+d\rho)\log(\rho+d\rho) - \rho\log\rho\big) \\ &= -\int \left((\rho+d\rho)\left(\log\rho + \frac{d\rho}{\rho} - \frac{1}{2}\frac{d\rho^2}{\rho^2}\right) - \rho\log\rho\right) \\ &= -\int \left((\log\rho+1)d\rho + \frac{1}{2}\frac{d\rho^2}{\rho}\right) \\ &= -\int \left(\log\rho\, d\rho + \frac{1}{2}\frac{d\rho^2}{\rho}\right). \end{aligned} \tag{70}$$

(In the last step we used the fact that $\int d\rho = d\int\rho = d1 = 0$.) We will now substitute into here for $d\rho$ from (39); however, notice that the non-stochastic terms from that equation will only contribute linearly (since terms of order $dt\,dW$ and dt^2 are negligible), so their final contribution to dS will be the same as already deduced in connection to master equation (27) (c.f. (31)). We therefore need only calculate here the contribution to dS of the stochastic term in (39); that is of $d\rho = \sqrt{8k}(A-\langle A\rangle)\rho\,dW$. Substituting this into (70), and in the following step using the rule of Itô calculus $dW^2 = dt$:

$$\begin{aligned} dS &= -\int \left(\log\rho\left(\sqrt{8k}(A-\langle A\rangle)\rho\,dW\right) + \frac{1}{2}\frac{\left(\sqrt{8k}(A-\langle A\rangle)\rho\,dW\right)^2}{\rho}\right) \\ &= -\sqrt{8k}\,dW\int(A-\langle A\rangle)\rho\log\rho - 4k dt\int(A-\langle A\rangle)^2\rho \\ &= -\sqrt{8k}\,dW\,\langle(A-\langle A\rangle)\log\rho\rangle - 4k\sigma_A^2 dt. \end{aligned} \tag{71}$$

This, together with the contribution (31) due to the non-stochastic terms from (39), gives us (40), as desired.

I.C Derivation of the hierarchy of equations (44)

Our objective is to derive the hierarchy of equations (44), which describes the contribution of the stochastic term in (39) to the evolution of the cumulants of $\rho(A;t)$ when $A = A(q,p)$ is not explicitly time-dependent. For brevity of notation we will omit the integration

measure $d^n q d^n p$ in integrals over phase space. Consider the cumulant-generating function for $\rho(A;t)$:

$$f(z;t) \triangleq \log\left\langle e^{zA}\right\rangle \triangleq \kappa_1(t)\frac{z}{1!} + \kappa_2(t)\frac{z^2}{2!} + \kappa_3(t)\frac{z^3}{3!} + \dots . \tag{72}$$

Let df denote the differential of this function with respect to time, and f' denote its derivative with respect to the dummy variable z. Expanding the differential of f to second order in $d\rho$:

$$\begin{aligned} df &= d\left(\log\int \rho\, e^{zA}\right) = \log\int (\rho + d\rho)e^{zA} - \log\int \rho e^{zA} \\ &= \left(\frac{\int d\rho\, e^{zA}}{\int \rho\, e^{zA}}\right) - \frac{1}{2}\left(\frac{\int d\rho\, e^{zA}}{\int \rho\, e^{zA}}\right)^2 . \end{aligned} \tag{73}$$

Substituting into here the stochastic term from (39) (that is $d\rho = \sqrt{8k}(A - \langle A\rangle)\rho\, dW$), and using the rule of Itô calculus $dW^2 = dt$:

$$\begin{aligned} df &= \sqrt{8k}\, dW\left(\frac{\int (A - \langle A\rangle)\rho\, e^{zA}}{\int \rho\, e^{zA}}\right) - 4k dt\left(\frac{\int (A - \langle A\rangle)\rho\, e^{zA}}{\int \rho\, e^{zA}}\right)^2 \\ &= \sqrt{8k}\, dW\left(f' - \langle A\rangle\right) - 4k dt\left(f' - \langle A\rangle\right)^2 . \end{aligned} \tag{74}$$

Writing f in terms of its cumulant expansion (72), and noting that $\langle A\rangle = \kappa_1$:

$$\begin{aligned} d\kappa_1\frac{z}{1!} + d\kappa_2\frac{z^2}{2!} + d\kappa_3\frac{z^3}{3!} + \dots = \sqrt{8k}\, dW\left(\kappa_2\frac{z}{1!} + \kappa_3\frac{z^2}{2!} + \kappa_4\frac{z^3}{3!} + \dots\right) \\ - 4k dt\left(\kappa_2\frac{z}{1!} + \kappa_3\frac{z^2}{2!} + \kappa_4\frac{z^3}{3!} + \dots\right)^2 . \end{aligned} \tag{75}$$

Expanding the square on the r.h.s. and equating coefficients of corresponding powers of z yields the hierarchy of equations (44), as desired.

Part II

An algorithm for locale navigation in rodents, with focus on the role of hippocampus[1]

Abstract

We propose a model regarding the computations taking place in the deliberative decision-making system of rodents, during wakefulness and sleep, with focus on the role of hippocampus (HPC). In this model, medial prefrontal cortex performs high-level planning, and then tasks HPC with fleshing out the details of the plan, as needed. We describe this planning task of HPC as an optimal control problem. Drawing insights from the powerful mathematics of optimal control theory, we provide a concrete algorithm by which HPC solves the problem; we point out the main sources of algorithmic complexity and how these shape the "computational life" of HPC; and we identify elements of the algorithm with prominent features of the neurophysiology of the system; such as the theta rhythm, the slow oscillation, spindle oscillations, sharp wave-ripples, θ-sequences, forward and reverse SWR-sequences, the formation and strengthening of episodic memories, and a need for two modes of operation—online and offline. The model may also provide novel insights into memory consolidation during sleep. In this way, the model offers a paradigm capable of accommodating a wide range of observed phenomena, and makes many novel testable predictions. Testing some of these predictions is the topic of ongoing work, and lies beyond the scope of this thesis.

II.1 Introduction

In the mammalian brain, the hippocampus (HPC) is a major structure located bilaterally in the medial temporal lobe. Being one of the most extensibly studied brain structures, much has been learned about HPC anatomy and physiology, but many questions remain unanswered, specially regarding its function. Two influential long-standing hypotheses implicate HPC, on the one hand, (i) as the locus of temporary episodic memory traces, which get consolidated into long-term memory during sleep [47]; and on the other hand, (ii) as the locus of the *cognitive map*, believed to underlie much of our spatial memory and our sense of spatial orientation [48]. Hypothesis (i) stemmed from observing particular forms of memory deficits in lesion studies. Hypothesis (ii) stemmed from the discovery that individual principal neurons of the HPC—now called *place cells*—become active only in some spatial environments, and not in others, and then only when the animal is in a particular spatially-localized region of the environment—the neuron's *place field*. In this

[1] This Part is adapted from a paper provisionally titled "A model of hippocampal sequences as solutions to the equations of optimal control", co-authored with Zhe S. Chen and Matthew A. Wilson, not yet submitted for publication.

Box 1 | Levels of analysis in brain sciences

As emphasized by Marr & Poggio [60], to understand a complex information-processing system, such as the deliberative system, it is useful to target our models at one of several levels of abstraction. In order of increasing abstraction these levels are usually taken to be: (i) mechanistic or implementational, (ii) algorithmic or representational, and (iii) computational or functional. The idea is that, as a first approximation, models at different levels can be considered in isolation, greatly simplifying the task of understanding. Furthermore, upon closer inspection, models at different levels can inform each another in useful ways. For example, in the descending direction: a well-established hypothesis regarding function will inform candidate algorithmic-level models, in the obvious sense that algorithms invented by people are always designed with functions in mind. In the ascending direction: limitations due to the types of algorithms that can be implemented by neural architecture will shape the computations and function of the system.

way, the neural activity of the population carries information that identifies the animal's current environment and location within that environment.

Attempts to synthesize the two hypothesis followed [49, 50]. However, the panorama has expanded considerably over the past two decades, largely due to two types of development: (i) the discovery of various forms of structure on short time-scales in the neural population activity of HPC (Section II.2.1). And (ii) a growing body of data regarding the interactions between HPC and other brain regions during wake and sleep (Section II.2.2). These data seem to outline a well-defined functional circuit, of which HPC is only one component. Indeed, recent proposals for synthesis situate HPC, along with prefrontal cortex (PFC), thalamus, and other structures, as part of a wider brain system believed to enable flexible behavior—primarily, but not exclusively, spatial behavior—in the face of changing contingencies [51–57]. This system has been referred to as the *deliberative* decision-making system, and has been contrasted with the *procedural* decision-making system involved in the formation of habits [58, 59].

In this Part, we begin by proposing our own synthesis of various data on the deliberative system, in the form of a system-wide algorithmic-level model (cf. Box 1). This model is meant to reproduce the function mentioned above—flexible behavior in the face of changing contingencies—while assigning tasks to individual brain structures, and interactions between structures, that are consistent with available data. Our model recapitulates a number of elements from previous syntheses of the PFC–HPC circuit; primarily from those of Eichenbaum [57], Penagos *et al.* [56], and Redish [55]. Put briefly, our algorithm involves two modes of operation: online and offline, and describes a hierarchical planning strategy in which medial PFC (mPFC) performs high-level planning, while HPC fleshes out the lower-level details of the plan. In spatial tasks, the latter computation takes the form of spatial trajectories, which appear in HPC as one of several types of "hippocampal sequences" (Section II.2.1), depending on the mode in which the algorithm is operating. When the algorithm operates in offline mode, new episodic memories are created, or existing ones are strengthened, after each calculation by HPC.

This system-wide algorithmic-level model ascribes to HPC a well-defined task that remains the same throughout both modes of operation. Namely: HPC receives as input initial and terminal states (two successive steps of the high-level plan) within a given context, and it must compute an efficient low-level plan to get from the initial to the terminal state. We formalize this planning task as an optimal control problem, arriving in this way at a computational-level model of HPC.

Drawing on the theory of optimal control, we go on to propose a concrete algorithm by which HPC solves its assigned task. We point out the main sources of complexity in this algorithm, and how these shape the "computational life" of HPC. We identify elements of the algorithm with prominent features of the neurophysiology of the system; such as the theta

rhythm, the slow oscillation, spindle oscillations, sharp wave-ripples, θ-sequences, forward and reverse SWR-sequences, the formation and strengthening of episodic memories, and a need for two modes of operation—online and offline.

The rest of this Part is organized as follows. Section II.2 reviews relevant background material; primarily on the neurophysiology of the deliberative system during wake and sleep, but also on established ideas about navigation and hierarchical planning. Section II.3 develops our three proposed models (together, "our model"); beginning with our system-wide algorithmic model (Section II.3.1), then our computational model of HPC (Section II.3.2), and finally our algorithmic model of HPC (Section II.3.3). Additional details and supporting evidence are provided in Boxes 2 and 3. Section II.3.4 points out potential insights into memory consolidation during sleep. In the discussion, Section II.4, we first give a brief summary of our model, as developed throughout the preceding sections. Then we point out our model's relationship to previous models in the literature. We take the time there to provide an in-depth comparison with two influential alternatives which also ascribe to HPC an active role in decision-making: the successor representation [61, 62], and the model by Mattar & Daw [63]. In Section II.4.3 we discuss many predictions and matters of interpretation of our model. In Section II.4.4 we pause to acknowledge the many limitations of scope of our model, and a few points of tension with empirical observations, which we have been able to identify so far. We close in Section II.4.5 pointing out the avenues for future research suggested by this work.

II.2 Background

II.2.1 Compressed sequences in the two states of hippocampus

Two states, or modes of operation, have been clearly identified in HPC; each associated with a distinct pattern of neural population activity, and waves of electrical activity in the local field potential (LFP). These states are named after the LFP patterns associated with them: *theta* and *large-amplitude irregular activity* (LIA). (Refer to Figure 1.) Theta state is the "online mode"; it occurs when the animal is engaged in active navigation and decision-making, as well as during REM sleep (i.e. dreaming) [64]. It is characterized by the presence of a persistent strong *theta rhythm* (~8Hz in rodents) in the LFP. The awake theta state is when HPC represents the current position of the animal and place fields can be observed. Interestingly, in this state not only the current position is represented, but also future positions immediately ahead; in such a way that the representation sweeps ahead of the animal once per theta oscillation, forming *θ-sequences* [65] (Figure 1b,c). In contrast, LIA state is the "offline mode"; it occurs during quiet wakefulness, such as during consummatory behavior, as well as during slow-wave (i.e. dreamless) sleep [66]. It is characterized by periods of low amplitude irregular fluctuations in the LFP, interrupted by brief large-amplitude deflections (*sharp waves*, 50–150ms) containing high-frequency oscillations (*ripples*, ~200Hz). These *sharp wave-ripples* (SWRs) correspond to intense bursts of activity in the population of neurons, which are not random, but organized, and often represent segments of local or remote trajectories; dubbed *SWR-sequences* [66, 67] (aka *replay*,[2] Figure 1d,e). Both kinds of sequences, θ- and SWR-, represent experience in a temporally-compressed way, with compression factors of 5–20x, so that relatively long stretches of trajectories (in rats: $\lesssim$50cm for θ-sequences [68]; $\lesssim$5m for SWR-sequences [69]) can be represented in a fraction of a second. Three notable differences worth highlighting between the two types of sequences are as follows. (i) Unlike awake θ-sequences which always begin near the location of the animal (Figure 1b,c), SWR-sequences may begin at remote locations [69] (Figure 1e), and even in other spatial contexts [70]. (ii) Unlike θ-sequences

[2] I favor "SWR-sequence", being a name less committed to a particular interpretation of the phenomenon.

which always represent experience in a "forward" fashion [71] (Figure 1b,c), SWR-sequences may represent experience both "forward" as well as "backward through time"; dubbed *forward* and *reverse SWR-sequences* [72] (Figure 1d,e). (iii) Unlike θ-sequences which always happen in isolation once per theta cycle (Figure 1b,c), SWR-sequences can be "chained together" within *SWR trains* (aka *multi-ripple bursts*, including up to eight SWRs), building up a single extended (forward or reverse) sequence out of multiple "atomic", or simple, SWR-sequences [69] (Figure 1e).

Concerning the possible functions of hippocampal sequences, partially complementary, partially conflicting hypotheses implicate them in episodic memory formation [77–81]; consolidation into long-term memory [52, 53, 75, 81–90]; maintaining, updating, or augmenting the cognitive map (including the discovery of generalizable cortical schemas) [53, 56, 80, 83, 90–92]; memory retrieval to guide behavior [75, 93, 94]; back-propagation of value for model-based reinforcement learning [62, 63, 72, 95–99]; and mental exploration for deliberative planning [55, 92, 95, 100–106]. As can be seen, many interesting ideas regarding the function of hippocampal sequences are being proposed, but a compelling paradigm is yet to emerge as dominant.

Figure 1: (Next page.) **Compressed sequences in the two states of hippocampus.** Left side of figure dedicated to θ-sequences; right side to SWR-sequences. (**a**) Bilateral LFP recorded in rat HPC, showing the signature 8Hz theta oscillation of the theta state ("walk"), and SWRs of the LIA state ("still"). Individual SWRs last 50–150ms, and can often occur one after another in short proximity; called SWR trains (aka multi-ripple bursts). (**b**) θ-sequences, which occur in the theta state, sweep up to $\lesssim$50cm ahead of the animal once per theta cycle (LFP shown at top), corresponding to time-compressions of $\lesssim$10x compared to real-time experience. The sequences are visible in the Bayesian decoder (bottom; color scheme is posterior probability) and even in the sorted raster plot (middle). (**c**) Early during learning, θ-sequences seem to flexibly explore future possibilities at decision points. Here we see a rat performing an alternation task on an unfamiliar M-maze (left). An example trial is illustrated in green. As the rat approaches the choice point, θ-sequences are seen to explore each of the two choices in alternation (right). (**d**) SWRs occur in the LIA state, when the rat is stationary or unengaged with the task. They are accompanied by intense bursts of activity which often play out either a forward or a reverse SWR-sequence (here visible in the sorted raster). Some SWR-sequences span across entire SWR trains that can last up to half a second. (**e**) By simultaneously decoding position and direction of motion one can reliably distinguish forward from reverse SWR-sequences. By definition, forward sequences advance in the direction of the decoded direction of motion, while reverse sequences advance opposite to it. Top: five SWR-sequences visualized in the Bayesian decoder (color saturation indicates posterior probability; blue (positive heading) and red (negative heading) indicate decoded direction of motion). Solid triangle indicates location of animal at time of event. Each sequence shown spans a SWR train, as can be seen from the multiple successive peaks in the multiunit activity (MUA, middle). Most SWR-sequences are neatly classified as either forward or reverse, but a fraction of them switch along the way from one class to another (example in top right; statistics in bottom left); so-called *mixed SWR-sequences*. Average SWR-sequence speed is ~8m/s (bottom right), corresponding to time-compressions of ~20x. Panel **a** is adapted from Buzsáki [73]; **b** from Feng *et al.* [68]; **c** from Kay *et al.* [74]; **d** from Carr *et al.* [75] (originally appeared in Diba & Buzsáki [76]); and **e** from Davidson *et al.* [69].

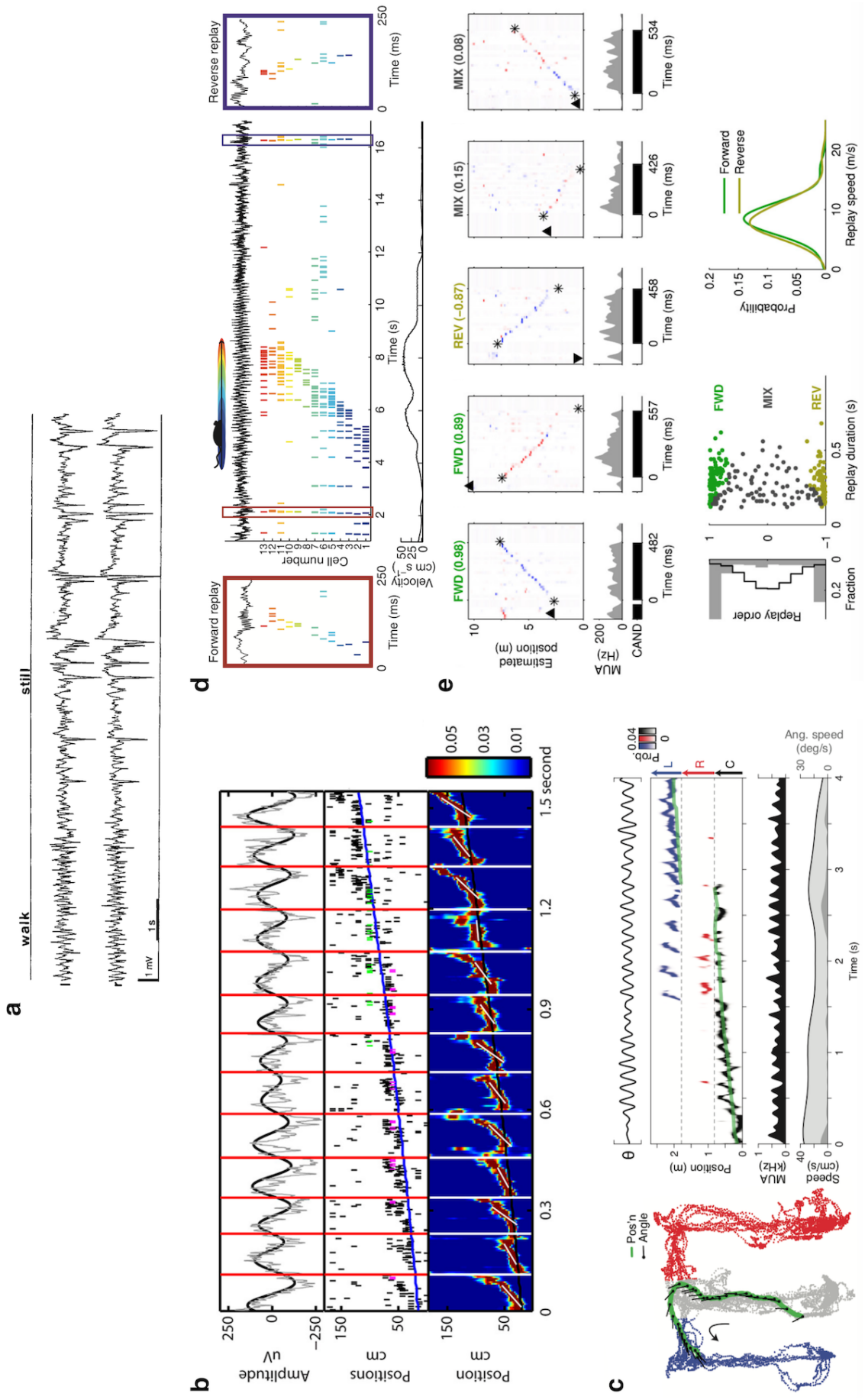


Figure 1: (Rotated. Caption previous page.)

II.2.2 PFC–HPC interactions during wake and sleep

Crossed-lesions studies provide robust evidence that mPFC and HPC support memory and decision-making via an ipsilateral pathway [107–111]. Moreover, the two states of HPC just discussed, theta and LIA, are observed to correspond to distinct states of mPFC. Indeed, distinct forms of coupled nested oscillations and oscillatory synchrony can be observed between the two regions during either state, as follows.

Wake. When the animal is engaged in navigation and memory tasks, the theta rhythm seen in HPC is also present in mPFC. The rhythms in these two regions are strongly synchronized [52, 112–120], and the strength of this coordination correlates with behavioral performance [121–125].

Moreover, studies that analyzed this interaction during successive stages of contextual memory tasks found that context cueing involves flow of information from HPC to mPFC, whereas context-appropriate decision-making involves flow of information from mPFC to HPC [120, 126]. Extending these findings, O'Neill *et al.* [124] showed that mPFC synchrony with ventral HPC (vHPC) supports performance even when the influence of dorsal HPC (dHPC) is removed; and Adhikari *et al.* [127] found that vHPC synchronized with, and led, mPFC in anxiety-inducing environments. Eichenbaum [57] has incorporated these and other findings into a model in which vHPC informs mPFC of task context, and then mPFC performs top-down control to gate the extraction of relevant information from dHPC to guide behavior. These ideas will form part of our model in the coming sections.

Sleep and quiet wakefulness. Refer to Figure 2. During slow-wave sleep, while HPC is in LIA state, cortex is engaged in the eponymous *slow oscillation* (SO, $\lesssim$1Hz) [128, 129]. The SO consists of alternations of suppressed (*down*) and elevated (*up*) neural activity across cortex, thalamus and HPC. Of special interest to us will be the *spindle oscillation* (~12Hz), which can be observed in cortex, typically near the beginning of each up sate [129]. Hippocampal SWRs also occur during the up state [129]. When SWRs occur in a train (aka multi-ripple burst), the individual SWRs tend to be phase-locked to individual cycles of the cortical spindle oscillation [130, 131]. As noted by Penagos *et al.* [56], the nested nature of these oscillations suggests a means by which complex computations can be decomposed into elementary operations across brain regions. This idea will also form part of our model in the coming sections.

Of note, hippocampal SWRs coordinate with PFC activity also during the awake resting state [93], suggesting similarities between the system's operation during slow-wave sleep and quiet wakefulness [132].

II.2.3 Five strategies of rodent navigation

It is useful to distinguish five strategies which an animal can employ for goal-directed navigation. Quoting from Redish *et al.* [49], these are

- *Random navigation.* If the animal has no information about the location of the goal, it must search randomly for it.
- *Taxon navigation.* The animal can find a cue toward which it can always run. For example, if the goal is visible, it can simply "run toward the goal".
- *Praxic navigation.* The animal can execute a constant motor program. For example, if the animal always starts at the same location, in the same orientation, and the goal is never moved, it can use praxic navigation to reach the goal.

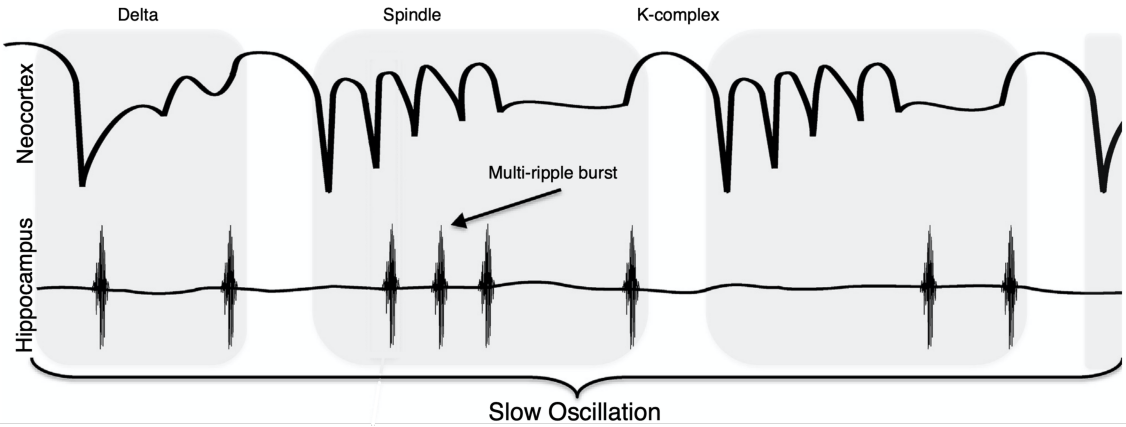


Figure 2: **Nested sleep oscillations in cortex and hippocampus.** Top trace, typical cortical LFP (low-pass filtered at ~20Hz), and bottom trace, typical hippocampal LFP (band-pass filtered at ~100–250Hz), during slow-wave sleep. The eponymous slow oscillation (SO, $\lesssim$1Hz) is seen in the cortical LFP (top). A large-amplitude biphasic wave known as *K-complex* marks the down state, during which neural activity is suppressed across cortex and HPC. During the up state (gray boxes), while neural activity is elevated, the spindle oscillation (~12Hz) can be observed in cortex, typically following the K-complex. *Delta waves* ($\lesssim$4Hz) may also be present during various phases of the cortical SO, but we will not consider them in this paper. Hippocampal SWRs (bottom, seen only as ripples due to filtering) occur during the up state. SWR trains (aka multi-ripple bursts) tend to be phase-locked to the cortical spindle oscillation. Figure adapted from Penagos *et al.* [56].

- *Route navigation.* The animal can learn to associate a direction with each sensory view. In more complex mazes, this entails planning a sequence of subgoals. For example, many early navigation tasks used complex mazes that consisted of sequences of T-junctions. Route navigation can be thought of as chaining sequences of taxon and praxic substrategies.

- *Locale navigation.* The animal can learn the location of the goal relative to a constellation of cues. It can learn a map on which the location of the goal is known. If it knows both its own location and the location of the goal in the same coordinate system, then it can plan a path from one to the other.

While the brain is capable of employing all of these strategies in combination, our concern in this paper will be exclusively with the latter strategy. Specifically: we address a part of the algorithm by which the deliberative system performs locale navigation.

II.2.4 Hierarchical mapping and planning

For reference in the coming sections, Figure 3 illustrates a generic scheme for two-stage hierarchical mapping and planning. First a low-level map of the state-space is built, and a high-level map is abstracted from it. With these maps at hand, the problem of planning a low-cost route from any start location ($\mathbf{S}_0$) to any goal location ($\mathbf{S}_g$) can be solved efficiently in a hierarchical fashion: first a high-level plan $\mathbf{S}_0 \to \mathbf{S}_1 \to \cdots \to \mathbf{S}_g$ is computed using the high-level map. Note that since the high-level map is, by construction, quite simple (e.g. a graph with relatively few nodes and edges), classical planning algorithms are adequate for this stage (e.g. A* search [133]). Next, each step of the high-level plan is refined using the low-level map. Notice that the high-level plan provides boundary conditions ($\mathbf{S}_i$, $\mathbf{S}_{i+1}$) for each stage of low-level planning. Since these boundary points are not too far apart from

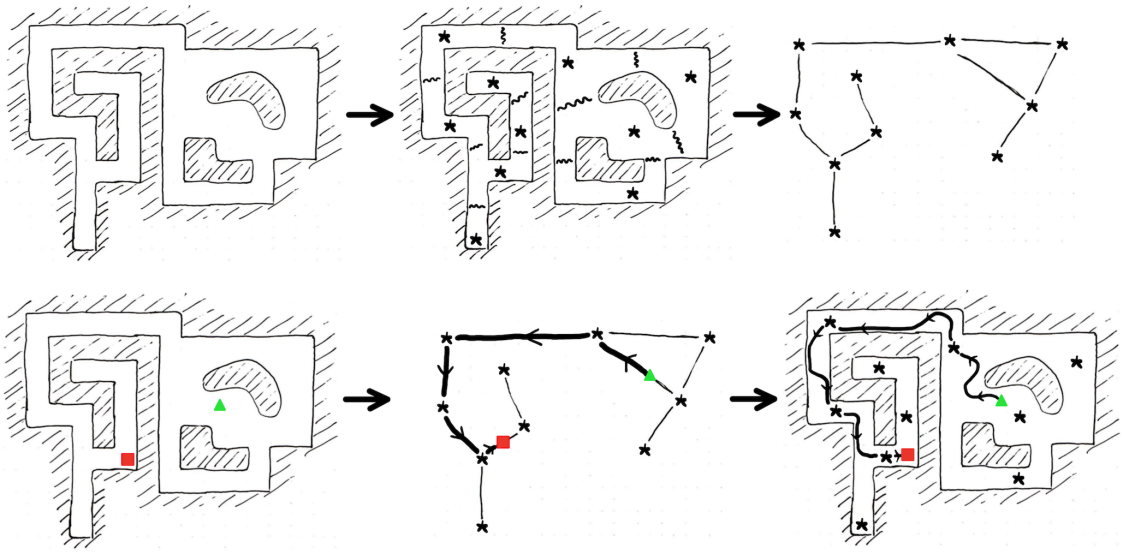


Figure 3: **Generic scheme for two-stage hierarchical mapping and planning.** Top: a high-level map (right) is abstracted from a low-level map (left). Bottom: Once the low- and high-level maps have been learned, efficient and flexible planning is carried out hierarchically from the top down.

each other, efficient continuum techniques (e.g. based on gradient descent) can be invoked during low-level planning, with little risk of getting stuck in bad local minima.

II.3 Modeling results

II.3.1 An algorithmic model of the deliberative system

Consistent with recent models of the deliberative system [51–57], we identify mPFC and HPC as key components of the system. Of HPC we further distinguish its dorsal (dHPC) and ventral (vHPC) regions. We follow Eichenbaum [57] in identifying vHPC as encoding task-relevant context; dHPC as encoding the animal's state within that context; and mPFC as encoding context-relevant rules, as well as exerting top-down control to gate task-relevant information at decision time. Our proposal is that the deliberative system enables flexible decision-making by implementing a two-stage hierarchical mapping and planning scheme, like that outlined in Section II.2.4. Specifically, we identify the cognitive map in dHPC with the low-level map, and the task rules of mPFC with the high-level map, in the hierarchical scheme of Figure 3. Following Eichenbaum [57], we propose that mPFC is informed of task context by vHPC, and is informed of the animal's state, $\mathbf{S}_0$, within that context by dHPC. We suggest that mPFC makes use of context, external state $\mathbf{S}_0$, internal motivational state, and context-relevant rules, to flexibly determine a high-level goal, $\mathbf{S}_g$, as well as a strategy to get from $\mathbf{S}_0$ to $\mathbf{S}_g$. This strategy takes the form of a high-level plan $\mathbf{S}_0 \to \mathbf{S}_1 \to \cdots \to \mathbf{S}_g$, as in Figure 3 (bottom center). Classical planning algorithms may be adequate for this computation, as mentioned in Section II.2.4. Next, the details of this plan need to be fleshed out before it can be acted upon, as in Figure 3 (bottom right).

At this point we distinguish two modes of operation of the system: *online* and *offline*. Online mode corresponds to the theta state of HPC and mPFC, and is the mode of operation whenever the animal is actively engaged in behavior. Offline mode corresponds to the LIA state of HPC, during which the slow oscillation can be seen in mPFC; it is the mode of operation during slow-wave sleep and quiet wakefulness. Online mode is characterized

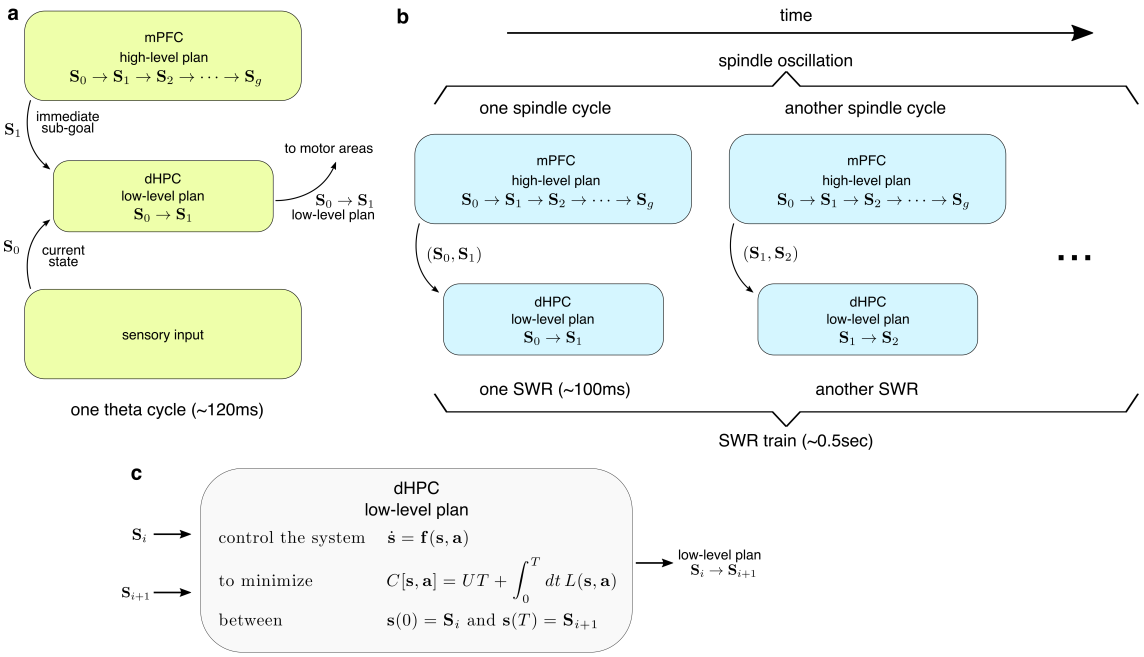


Figure 4: **An algorithmic model of the deliberative system, containing a computational model of dHPC.** (**a**) In online mode, mPFC computes a high-level plan appropriate for the current task context, external state $\mathbf{S}_0$, and internal motivations. The task of dHPC is to quickly refine the first step of this plan, and output the resulting low-level plan as a candidate for execution. This task of dHPC repeats on a duty cycle of ~120ms, giving rise to the characteristic 8Hz theta oscillation. (**b**) In offline mode, mPFC sustains more prolonged conversations with dHPC. In each exchange (50–150ms), mPFC prompts dHPC with the next step of its high-level plan, and dHPC tries to compute the corresponding low-level plan. The whole conversation takes place over the course of a cortical up state; it appears in mPFC as the spindle oscillation, and in dHPC as a SWR train threaded by a single extended SWR-sequence. This precomputation serves to ease the system's computational burden during subsequent online episodes, as detailed in Section II.3.3. (**c**) The computational task of dHPC remains the same in both modes of operation: we formalize it as an optimal control problem.

by an urgency to output actionable plans to guide behavior in real time. Hence, while in this mode, the system focuses its resources on refining only the first step of its high-level plan, ($\mathbf{S}_0$, $\mathbf{S}_1$)—the only step that is imminently needed. To do this, we propose that the system executes Algorithm 1, also illustrated in Figure 4a.

Algorithm 1. Online mode routine

1: `while` in online mode `do`: ▷ loop iterates at ~8Hz; theta rhythm
2: Sensory afferents pass current state, $\mathbf{S}_0$, to dHPC
3: mPFC passes first step of high-level plan, $\mathbf{S}_1$, to dHPC
4: dHPC computes efficient trajectory $\mathbf{S}_0 \to \mathbf{S}_1$; a low-level plan ▷ a θ-sequence
5: Resulting low-level plan is output to motor areas as candidate for execution

In the rodent, the `do while` loop in line 1 of this algorithm iterates on a duty cycle of ~120ms, giving rise to the synchronous PFC–HPC theta rhythm. Line 4 of this routine manifests in dHPC as a θ-sequence. The details of this step will be further fleshed out in Section II.3.3. Note that in each cycle of this routine, $\mathbf{S}_0$ stands for the current state of the animal, which changes gradually from iteration to iteration in accordance with the animal's displacement; and $\mathbf{S}_1$ stands for the first step of the high-level plan that still lies ahead.

Offline mode is characterized by a reduced urgency to output actionable plans. Thus, the system can dedicate itself to more lengthy computations; specifically, to precomputing answers that will ease its burden during subsequent online episodes. To do this, we propose that the system executes Algorithm 2, also illustrated in Figure 4b.

Algorithm 2. Offline mode routine

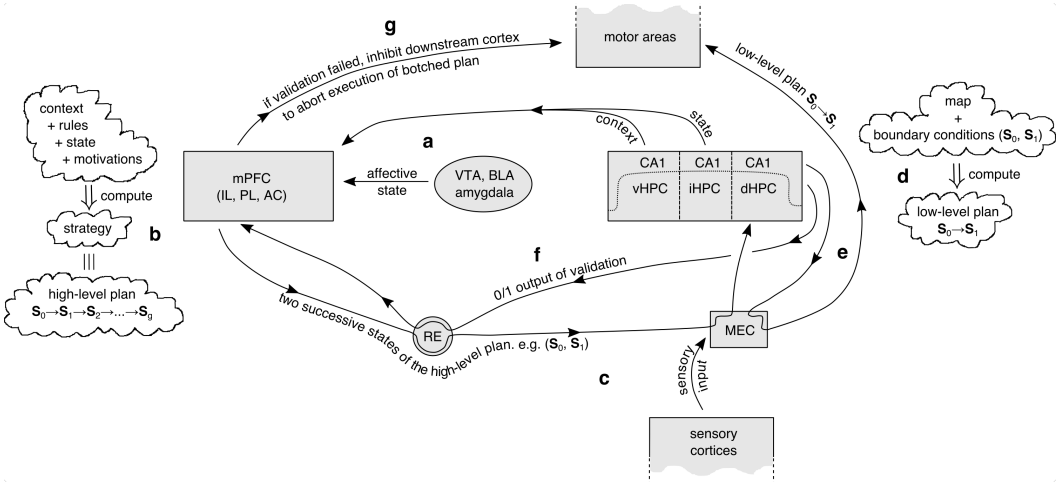
```
1: while in offline mode do:                                   ▷ loop iterates at ≲1Hz; slow oscillation
2:     mPFC computes a high-level plan S_0 → S_1 → ··· → S_g
3:     for i ∈ {0, ..., g − 1}:                                ▷ loop iterates at ~12Hz; spindle oscillation
4:         mPFC passes step (S_i, S_{i+1}) of high-level plan to dHPC
5:         dHPC computes efficient trajectory S_i → S_{i+1}; a low-level plan   ▷ a SWR-sequence
6:         if dHPC's computation fails then:
7:             break
```

The `do while` loop in line 1 of this algorithm corresponds to the PFC-thalamo-HPC slow oscillation, while the `for` loop in line 3 corresponds to the nested spindle oscillation (Figure 2). In Section II.3.3 we will have more to say about how mPFC chooses which plans to compute in line 2. Line 5 manifests as a SWR-sequence in dHPC. The details of this step will also be fleshed out in Section II.3.3. The `break` clause in lines 6 and 7 controls the length of the SWR train; if the break happens on the first iteration the result is a simple SWR. More details on this clause (such as what it means for dHPC's computation to "fail") are given in Box 2, and in Section II.3.3. In that Section we will also explain how the results of these offline computations are efficiently stored, and how they are used to ease the computational burden during subsequent online episodes.

In Box 2 we give a more detailed description of the model presented in this section, which includes also roles for entorhinal cortex, the thalamic nucleus reuniens, and affective regions such as the ventral tegmentum and amygdala. There we also summarize the literature supporting the various elements of the model.

Box 2 | An algorithmic model of the deliberative system; further details

In this Box we give a more detailed description of our model of the deliberative system presented in Section II.3.1 of the main text. This model is an elaboration of a model by Eichenbaum [57], incorporating also ideas from Redish [55] and Penagos *et al.* [56]. After describing our model we summarize the supporting evidence for its various elements.



Model description. (**a**) vHPC encodes task-relevant context, while dHPC encodes the animal's state within that context. From its output in CA1, via the direct HPC–mPFC pathway, HPC informs ipsilateral mPFC of context and state. mPFC also receives the internal affective state from amygdala and VTA. (**b**) mPFC itself encodes high-level context-dependent task rules. Drawing upon these four ingredients, mPFC flexibly computes an appropriate strategy, in the

form of a high-level plan $\mathbf{S}_0 \to \mathbf{S}_1 \to \cdots \to \mathbf{S}_g$ within the given context, leading from the given starting state $\mathbf{S}_0$, to a goal state $\mathbf{S}_g$ that would satisfy internal motivations, using valid transitions that follow the rules. This mPFC computation is done on a simplified high-level map, so that classical planning techniques (e.g. A*) may be tractable. Before it can be a candidate for execution, this high-level plan must be refined. (**c**) To refine its high-level plan, mPFC pings ipsilateral dHPC with the first two steps of its plan $(\mathbf{S}_0, \mathbf{S}_1)$. This information follows the pathway through RE and then MEC. During "online mode" (see below) this mPFC efferent is integrated with sensory information in MEC to improve the estimate of $\mathbf{S}_0$; during "offline mode" (see below) sensory input is cutoff. In any case, dHPC receives two successive states of the mPFC high-level plan. (**d**) dHPC encodes a continuous spatial map for the context at hand. The two states it receives from mPFC define two points on this map, not too far apart from one another. The task of dHPC is to compute a low-cost continuous trajectory—a low-level plan $\mathbf{S}_0 \to \mathbf{S}_1$—that admits these two points as boundaries. This computation manifests as a HPC sequence. (See main text, Section II.3.3, for how HPC might do this.) At this point we distinguish two modes of operation: *online* and *offline*. (**e**) During goal-directed behavior, the system operates in online mode: low-level plans produced by dHPC are broadcast from dorsal CA1 to motor-related areas as candidates for execution. (**f**) However, it is possible for the low-level plan to have failed to reach the desired sub-goal, $\mathbf{S}_1$, leading instead to some other location $\mathbf{S}'_1$ (see main text, Section II.3.3). The plan was broadcast to motor areas nonetheless (step **e**), because pausing to verify would introduce delays that would be in tension with the demands of real-time behavior. To guard against the possibility of a botched plan by HPC: simultaneously with **e**, there is a validation step at dorsal CA1 to check whether $\mathbf{S}'_1$ is close enough to $\mathbf{S}_1$ within some margin of error. The result of this validation (one bit of information) is passed to ipsilateral mPFC via RE. (**g**) If the validation failed, mPFC responds by inhibiting motor areas, preventing the animal from acting on HPC's botched plan. In any case, arrival of the validation signal to mPFC prompts mPFC to ping ipsilateral dHPC with $(\mathbf{S}_0, \mathbf{S}_1)$ once more (step **c** again). Steps **c**–**g** continue to iterate in this way, on a duty cycle of ~120ms in the rodent (the theta rhythm); in each cycle, $\mathbf{S}_0$ stands for the current state of the animal, which changes gradually from iteration to iteration according to the animal's displacement; and $\mathbf{S}_1$ stands for the first step of the high-level plan that still lies ahead. In this way the animal performs real-time flexible planning while behaving. The other mode of operation, offline mode, takes place during quiet wakefulness and slow-wave sleep. In this mode, low-level plans produced by HPC are not broadcast to downstream cortex (there are no steps **e**, **g**). This mode is characterized by a reduced urgency to output actionable plans. Thus, the system can dedicate itself to more lengthy computations; specifically, to precomputing answers that will ease its burden during subsequent online episodes, as follows. Now if the validation step in dorsal CA1 succeeds, mPFC (upon hearing of the success, step **f**) replies by prompting dHPC with the *next* two steps of its high-level plan (step **c** again, but now with $(\mathbf{S}_1, \mathbf{S}_2)$ in place of $(\mathbf{S}_0, \mathbf{S}_1)$); to which dHPC responds by again computing the corresponding low-level plan $(\mathbf{S}_1 \to \mathbf{S}_2)$. Assuming rest is not interrupted, this back-and-forth continues to iterate over successive steps of the high-level plan, until one of HPC's low-level plans fails to validate, or until the entire mPFC high-level plan is fleshed out. Each back-and-forth takes 50–150ms; in HPC each manifests as a SWR; in mPFC as a cycle of the spindle oscillation. The complete exchange appears in HPC as a SWR train ($\lesssim$0.5sec), threaded by a single extended SWR-sequence. Section II.3.3 of the main text explains how the results of these offline computations are efficiently stored, and how they are used during subsequent online episodes. Abbreviations: mPFC = medial prefrontal cortex; IL = infralimbic cortex; PL = prelimbic cortex; AC = anterior cingulate cortex; MEC = medial entorhinal cortex; vHPC = ventral hippocampus; iHPC = intermediate hippocampus; dHPC = dorsal hippocampus; RE = thalamic nucleus reuniens; VTA = ventral tegmental area; BLA = basolateral amygdala.

Supporting evidence. The model presented in this Box is an elaboration of a model of the mPFC-thalamo-HPC circuit by Eichenbaum [57], incorporating also ideas from Redish [55] and Penagos *et al.* [56]. Therefore much of the same supporting evidence discussed in those papers can be invoked to support our model. Here we briefly recapitulate some of that evidence.

HPC function: It is well established that HPC plays a role in tasks that demand remember-

ing events in the spatial context in which they occurred [134–140]. In agreement with this, many studies have reported that hippocampal neurons remap between tasks, forming novel allocentric spatial codes that allow placing memories in context [141–148]. In addition, it has recently been found that there is a topography to HPC memory representations [149]: namely, while neurons in dHPC of rodents encode highly specific locations, neurons in vHPC encode large areas of space [150, 151] and distinguish events that occur in different spatial contexts [152–155]. Correspondingly, lesioning vHPC, but not dHPC, attenuated the acquisition and expression of contextual fear conditioning, whereas lesioning dHPC, but not vHPC, dramatically impaired performance on a spatial working memory task [156]. *mPFC function:* Much converging evidence indicates that PFC contributes to decision-making by top-down control of memory processing [53, 157–162]. In rats [163–168] and in monkeys [169–172], mPFC has been implicated in the selection and maintenance of "task sets", also called "task rules" or "options"—extended, context-specific sequences of behavior, directed toward particular goals, possibly as part of a hierarchical decision-making scheme [173, 174], allowing for the flexible switching between strategies in various tasks [175–185]. *Affective inputs to mPFC:* The mPFC is strongly interconnected with the BLA of the amygdala and with the VTA of the midbrain; and it has been proposed that mPFC is ideally positioned to integrate current and past information with its affective qualities in order to guide decision-making [186, 187]. *Direct HPC→mPFC pathways:* mPFC and HPC are known to be strongly connected by a few direct- and several indirect pathways. Two well-known direct pathways consist of monosynaptic projections from area CA1 of the vHPC and iHPC broadly to all layers of mPFC [187–189]. Since vHPC encodes spatial context, while iHPC is closer to the part of HPC which precisely encodes location, these two direct pathways may respectively inform mPFC of the task context and of the animal's specific location within that context. Consistent with this hypothesis, optogenetic inactivation of vHPC terminals in mPFC during context cueing, but not during a post-cueing delay period or at decision time, was found to impair spatial working memory [125]. *mPFC–HPC interactions during wake:* Refer to main text, Section II.2.2.

Interim summary #1: The above studies all provide converging evidence for the coding and computational roles ascribed by our model to HPC and mPFC in **a** and **b**, and for the pathway and content of their communication in **a**. These parts of our model are essentially the same as Eichenbaum [57]. We will show in the main text (Section II.3.3) how our model's proposed computational role for dHPC in **d** is consistent with prominent features of the neurophysiology of HPC, such as θ-sequences and forward and reverse SWR-sequences.

mPFC↔HPC pathways through RE: One of the most anatomically-salient indirect connections between mPFC and HPC involves thalamus as the intermediary [120, 186, 190, 191]. This pathway includes bidirectional connections between all mPFC areas and RE; and bidirectional connections between RE and HPC area CA1 throughout its dorsal-ventral extent, as well as between RE and entorhinal cortex [186, 192–195] (in particular MEC, which is known for its role in path integration [196–198]). In turn, entorhinal cortex has projections to all areas of HPC [155, 199–201]. Regarding the question of whether these connections (mPFC↔RE↔HPC) are functional, and if so, what their function might be: as mentioned in Section II.2.2 of the main text, context-appropriate behavior involves flow of information from mPFC to HPC at decision time [120, 126]. Specifically, Hallock *et al.* [120] found that transient inactivation of the mPFC↔RE↔HPC pathway by muscimol infusion in RE selectively disrupted (i) performance on a spatial working memory task, (ii) mPFC–dHPC theta synchrony at decision time, and (iii) information flow at decision time from mPFC to dHPC. Consistent with this, Ito *et al.* [202] reported goal-dependent firing in mPFC, RE and dHPC of rodents in a spatial memory task; and found that optogenetic silencing of RE significantly reduced goal-dependent firing in dHPC—suggesting that goal information is conveyed from mPFC to dHPC through RE at decision time.

Interim summary #2: These studies provide converging support for our model's proposed pathways of bidirectional communication between mPFC and dHPC in **c** and **f**, as well as for the content of this communication proposed in **c**.

mPFC's inhibitory control over motor output: Our model's proposed role for mPFC in **g**—that it inhibits motor-related areas in order to veto execution of a pre-committed course of action—is consistent with many human studies on this topic [203–208]. (For comparing primate studies with our rodent model, note that rodent mPFC is arguably homologous to primate dorsolateral

PFC [209, 210].) *Relationship to vicarious trial and error (VTA) behavior:* Our model predicts that if failure to validate (step **f**) recurs, as might be expected to happen early during learning of a novel task or at difficult choice points, this can yield successive start-then-stop motor commands (steps **e** and **g**, repeating on successive theta cycles). As discussed by Redish [55], this aspect of the model is consistent with VTA behavior, and makes several further neurophysiological predictions concerning VTA that are in general agreement with experimental observations. *Two modes of operation, and mPFC–HPC interactions during sleep:* Refer to main text, Section II.2.2.

II.3.2 A computational model of dHPC

> After a while he calmed down and explained to me that not every place was good to sit or be on, and that within the confines of the porch there was one spot that was unique, a spot where I could be at my very best. It was my task to distinguish it from all the other places. The general pattern was that I had to "feel" all the possible spots that were accessible until I could determine without a doubt which was the right one.
>
> —Carlos Castañeda, The teachings of Don Juan: a Yaki way of knowledge [211]

Our model of the deliberative system in Section II.3.1 (and Box 2) ascribes a function to dHPC that remains the same throughout the online and offline modes of operation. Namely: dHPC receives as input initial and terminal states, $(\mathbf{S}_i, \mathbf{S}_{i+1})$, within a given context, or state-space, $\mathcal{S}$, and it must compute a detailed plan to get from the initial to the terminal state. This detailed plan takes the form of a continuous trajectory through $\mathcal{S}$ joining the two endpoints. If $\mathbf{S}_{i+1}$ is within line-of-sight from $\mathbf{S}_i$, and there are no obstacles in between, then taxon navigation (cf. Section II.2.3) seems to offer a straightforward solution to the problem. However, under ethological conditions, taking the straight path might put the rodent out in the open where it is an easy target for flying predators. Solving the problem under these conditions calls for a more sophisticated approach. One way to formulate the problem is to ascribe cost to regions of space (fear conditioning) according to how dangerous, or otherwise inconvenient, they are for the animal to traverse (see example in Figure 5). The computational task of HPC can then be cast as an optimal control problem (Figure 4c):

$$\text{control the system} \qquad \dot{\mathbf{s}}(t) = \mathbf{f}(\mathbf{s}(t), \mathbf{a}(t)) \tag{1a}$$

$$\text{to minimize} \qquad C[\mathbf{s}, \mathbf{a}] = UT + \int_0^T dt\, L(\mathbf{s}(t), \mathbf{a}(t)) \tag{1b}$$

$$\text{where} \qquad \mathbf{s}(0) = \mathbf{S}_i \text{ and } \mathbf{s}(T) = \mathbf{S}_{i+1} \text{ are fixed; while } T \geq 0 \text{ is free.} \tag{1c}$$

Our notation in this Part is that bold mathematical symbols denote vectors (more generally, points on a manifold); $\mathbf{s}(t)$ is the represented location of the animal (more generally, its *state*) at time t, which we take to lie in a continuous n-dimensional state-manifold, $\mathcal{S}$. (We have in mind typically $n = 1, 2$ or 3, corresponding, e.g., to a 1D track, a 2D maze or arena, or 3D physical space.) And $\mathbf{a}(t)$ is the *action* prescribed by the plan at time t, which we take to be chosen from a continuous m-dimensional action-manifold, $\mathcal{A}$. (We have in mind $m \leq n$, typically.) The differential equation (1a), the *state equation*, is meant to be a simplified model of the animal's motor system, capturing the effect of taking any action $\mathbf{a}$ while the animal is in any state $\mathbf{s}$. The function $\mathbf{f}$ in this equation is the *transition function*. In (1b), the *cost functional* $C[\mathbf{s}, \mathbf{a}]$ assigns a real number to each possible plan $\{\mathbf{s}(t), \mathbf{a}(t)\}_t$, as specified by the formula on the right. The integrand $L(\mathbf{s}, \mathbf{a})$ here is the *cost-rate* (its negative is the effective *reward-rate*), describing the rate at which cost accumulates, as function of state and action. The first part of (1c) recapitulates the

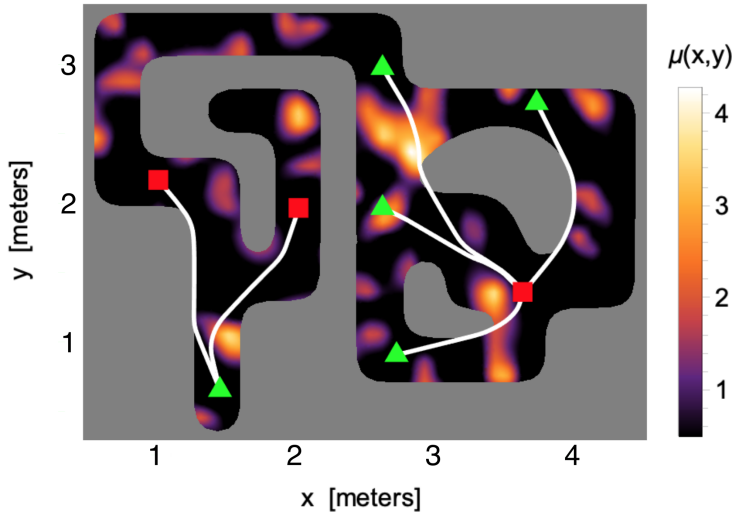


Figure 5: **Example of the task ascribed to dHPC by our computational model.** Example of the optimal control problem (1) using 2D state and action spaces ($n = m = 2$). Here we set $\mathbf{f}(\mathbf{s}, \mathbf{a}) \triangleq \mathbf{a}$, which corresponds to complete locomotive control; and $L(\mathbf{s}, \mathbf{a}) \triangleq \frac{\mu(\mathbf{s})}{2}\|\mathbf{a}\|^2$, which describes a cost to locomotion (e.g. metabolic cost, or risk of predation) that is quadratic in the control effort, with the coefficient of proportionality, $\mu(\mathbf{s})/2$, being location-dependent, as may be adequate for describing variations in terrain conditions (altitude, ruggedness), ambient conditions (heat, cold, wind), risk of predation, fear conditioning, etc. The figure defines the 2D state space $\mathcal{S}$ (gray marks inaccessible regions of the plane); as well as the "contextual fear" function $\mu(\mathbf{s})$ (heat map). The action space is $\mathcal{A} = \mathbb{R}^2$. Optimal trajectories are shown between a few example start (green triangles) and end (red squares) locations.

boundary conditions $(\mathbf{S}_i, \mathbf{S}_{i+1})$ dictated as input to dHPC. T is the time allotted by the plan for transitioning from $\mathbf{S}_i$ to $\mathbf{S}_{i+1}$, which we treat as a free parameter to be determined as part of the optimization. The term UT appearing in (1b) penalizes plans for how long they take. Here $U > 0$ is a fixed constant that can be understood as a measure of urgency; larger values of U more heavily incentivize brief travel times. We proceed in this paper without making any assumptions regarding the particular forms of the two functions $\mathbf{f}$ and L, or the value of U. Figure 5 illustrates the kind of problem we have in mind, and the kind of solutions defined by (1).

II.3.3 An algorithmic model of dHPC

Solving the optimal control problem (1) by brute-force search is exponentially slow (see Box 3), so it would not be a viable option for HPC. For fast methods we look to optimal control theory [46]. Generally speaking, two tractable approaches are known: the *Hamilton-Jacobi-Bellman* (HJB) *equation*, and *Pontryagin's maximum principle*. We describe both of these approaches in Box 3, where we also explain our reasons for believing that Pontryagin offers the better platform upon which to model the computations taking place in dHPC, given the task ascribed by our model above. In this section we flesh out this idea in the form of an algorithmic-level model of dHPC.

As discussed in Box 3, Pontryagin's principle provides a system of ordinary differential equations (ODEs) ((iia–iic) of Box 3) for the optimal plan that solves (1). The unknowns in these ODEs are $\mathbf{s}^\star(t) \in \mathcal{S}$, $\mathbf{a}^\star(t) \in \mathcal{A}$, and $\mathbf{p}^\star(t) \in \mathbb{R}^n$; respectively, the optimal state, action, and *co-state* trajectories. The co-state, $\mathbf{p}$, is a new "dummy variable" appearing

in Pontryagin's principle, which plays an auxiliary role in the calculation, and which has the same dimensionality, n, as the state $\mathbf{s}$. The ODEs must be satisfied throughout the duration of the plan. As explained in the Box, to determine a unique solution, a set of $2n$ boundary conditions must be satisfied (cf. (1c)):

$$\mathbf{s}^\star(0) = \mathbf{S}_i \quad \text{and} \quad \mathbf{s}^\star(T) = \mathbf{S}_{i+1}. \tag{2}$$

Notice, importantly, that these boundary conditions are split between the two endpoints of integration. We see that the computational task of dHPC, in our model, is reduced to solving a system of ODEs subject to split boundary conditions; a two-point boundary value problem (TPBVP).

The above provides the setup for understanding the "computational life" of dHPC in our model. Consider the following three key points. (i) If all boundary conditions could be given at the initial time $t = 0$, so that instead of a TPBVP we had an initial-value problem (IVP), then planning could easily be done at query time by a single forward sweep of numerical integration of the ODEs. (ii) But because the boundary conditions are split between the two endpoints of integration, the problem is more challenging, and dHPC is forced to rely on precomputation during offline periods. An efficient way to organize such precomputation is described in Box 3, and we make that idea a part of our model as follows. During online episodes, while the animal is engaged in behavior, we propose that the brain flags those states of the environment that are salient as either *goal locations* (states towards which the animal may often need to navigate) or *origin locations* (states from whence the animal may often need to depart). We think of these locations as hubs for navigation. During subsequent periods of rest—when there is time to spare for computation—PFC informs dHPC of the flagged states, and dHPC executes a "blindfolded forward shooting method" (forward SWR-sequences) at the flagged starting locations, and a "blindfolded reverse shooting method" (reverse SWR-sequences) at the flagged goal locations. (For a description of these shooting methods see Box 3.) The result is a collection of optimal plans leading from the flagged starting locations out to many places on the map, and from many places on the map in to the flagged goal locations. Now, this collection of plans would need to be committed to memory (episodic memories) for retrieval during subsequent online episodes; and there would need to be one such collection for each experienced context, $\mathcal{S}$. This would amount to a large library of possible plans, raising the question of how such a library could be stored efficiently. This leads to our third key point. (iii) Each time during rest that dHPC computes an optimal plan $\{\mathbf{s}^\star(t), \mathbf{p}^\star(t), \mathbf{a}^\star(t)\}_t$ between some initial and terminal states $(\mathbf{s}^\star(0), \mathbf{s}^\star(T))$, instead of storing the full plan, it suffices to store only the association

$$(\mathbf{s}^\star(0), \mathbf{s}^\star(T)) \mapsto (\mathbf{p}^\star(0), \mathbf{p}^\star(T)), \tag{3}$$

between the initial and terminal states and the initial and terminal co-states of the corresponding optimal plan. This requires drastically less memory; and yet $(\mathbf{s}^\star(0), \mathbf{p}^\star(0))$ is a seed from which the optimal plan can be quickly reconstructed at query time by a single forward sweep of numerical integration (a θ-sequence), because these data constitute a full set of initial conditions for Pontryagin's equations (cf. point (i)). In summary, what is bought by the precomputation is the conversion of the difficult TPBVP into the easy IVP, which can then be solved in real-time, at query time, whenever needed during subsequent online episodes; and further, these precomputations are done with a preference for optimal plans to, and from, salient hub locations on the map.

We make these ideas concrete in Algorithms 3 and 4, which are fleshed-out versions of the two algorithms from Section II.3.1.

Algorithm 3. Online mode routine; further details

1:	`while` in online mode `do`:	▷ loop iterates at ~8Hz; theta rhythm
2:	Sensory afferents pass current state, $\mathbf{S}_0$, to dHPC	
3:	mPFC passes first step of high-level plan, $\mathbf{S}_1$, to dHPC	
4:	`if` dHPC contains memory $(\mathbf{S}_0, \mathbf{S}_1) \mapsto (\mathbf{p}_0^\star, \mathbf{p}_1^\star)$ `then`:	
5:	dHPC does forward sweep of integration with ICs $(\mathbf{S}_0, \mathbf{p}_0^\star)$	▷ a goal-directed θ-sequence
6:	`else`:	
7:	dHPC does an iteration of forward shooting method from $\mathbf{S}_0$	▷ an exploratory θ-sequence
8:	Resulting low-level plan is output to motor areas as candidate for execution	

In line 4 of this algorithm, $(\mathbf{p}_0^\star, \mathbf{p}_1^\star)$ stands for the initial and terminal co-states corresponding to the optimal plan leading from $\mathbf{S}_0$ to $\mathbf{S}_1$, as in (3). (Similar comments apply to lines 8 and 21 of Algorithm 4, below.) The memory alluded to there is an episodic memory, whose formation is proposed to take place during offline periods (see below). In line 5 (and in lines 9 and 22 of Algorithm 4), "ICs" and "TCs" stand for "initial conditions" and "terminal conditions", respectively. As indicated, line 5 appears in dHPC as a goal-directed θ-sequence, as are observed in rodents when they have become proficient at a task [55]; while line 7 appears as an exploratory θ-sequence, as are observed in rodents during the early stages of learning [55, 74] (Figure 1c). Computationally, the purpose of a goal-directed θ-sequence is to recreate an imminently needed optimal trajectory, which has already been precomputed before, during rest. Notice that the forward sweep of integration required in this reconstruction is not only of very low algorithmic cost, but it is also an example of an *anytime algorithm*; an algorithm that returns a valid solution even if it is interrupted before it ends. This would explain why θ-sequences do not need to be as long as SWR-sequences. In contrast, the purpose of exploratory θ-sequences is as a last-ditch attempt to solve the challenging TPBVP, by executing the forward shooting method on the fly. This is likely to require several shooting attempts before yielding a solution passing near the required terminal condition—meanwhile leaving the animal "lost in thought", exhibiting vicarious trial and error behavior (see "Relationship to vicarious trial and error behavior", in Box 2) [55]. Somewhat similarly, Algorithm 4, below, stipulates conditions under which the system will produce SWR-sequences that are either goal-directed or exploratory; but also forward or reverse, and simple or trains. This algorithm also stipulates criteria for the formation and strengthening of episodic memories in dHPC. The purpose of these memories is for them to be used during subsequent online episodes, as just described.

Algorithm 4. Offline mode routine; further details

`while` in offline mode `do`: ▷ loop iterates at $\lesssim$1Hz; slow oscillation
  mPFC selects between doing forward (F) or reverse (R) type trial ▷ up-state of slow oscillation is dedicated to one of two trial types
  `if` an F-type trial `then`:
    mPFC selects a previously-flagged salient origin, $\mathbf{S}_0$
    mPFC computes a high-level plan efferent from $\mathbf{S}_0$: $\mathbf{S}_0 \to \mathbf{S}_1 \to \cdots \to \mathbf{S}_g$
    `for` $i = 0$ to $(g-1)$: ▷ loop iterates at ~12Hz; spindle oscillation
      mPFC passes step $(\mathbf{S}_i, \mathbf{S}_{i+1})$ of high-level plan to dHPC
      `if` dHPC contains memory $(\mathbf{S}_i, \mathbf{S}_{i+1}) \mapsto (\mathbf{p}^\star_i, \mathbf{p}^\star_{i+1})$ `then`:
        dHPC does forward sweep of integration with ICs $(\mathbf{S}_i, \mathbf{p}^\star_i) \Rightarrow$ optimal plan, $\{\mathbf{s}^\star(t), \mathbf{p}^\star(t)\}_t$, terminating near $\mathbf{S}_{i+1}$ ▷ a goal-directed forward SWR-sequence
        commit association $(\mathbf{s}^\star(0), \mathbf{s}^\star(T)) \mapsto (\mathbf{p}^\star(0), \mathbf{p}^\star(T))$ to memory in dHPC ▷ existing episodic memory strengthened
      `else` (memory not found):
        dHPC does an iteration of forward shooting method from $\mathbf{S}_i \Rightarrow$ optimal plan, $\{\mathbf{s}^\star(t), \mathbf{p}^\star(t)\}_t$; may not terminate near $\mathbf{S}_{i+1}$ ▷ an exploratory forward SWR-sequence
        commit association $(\mathbf{s}^\star(0), \mathbf{s}^\star(T)) \mapsto (\mathbf{p}^\star(0), \mathbf{p}^\star(T))$ to memory in dHPC ▷ new episodic memory formed
      `if` $\mathbf{s}^\star(T)$ is far from $\mathbf{S}_{i+1}$ `then`:
        `break` ▷ SWR train has derailed; terminate trial
  `else` (an R-type trial):
    mPFC selects a previously-flagged salient goal, $\mathbf{S}_g$
    mPFC computes a high-level plan afferent to $\mathbf{S}_g$: $\mathbf{S}_0 \to \mathbf{S}_1 \to \cdots \to \mathbf{S}_g$
    `for` $i = (g-1)$ to 0: ▷ loop iterates at ~12Hz; spindle oscillation
      mPFC passes step $(\mathbf{S}_i, \mathbf{S}_{i+1})$ of high-level plan to dHPC
      `if` dHPC contains memory $(\mathbf{S}_i, \mathbf{S}_{i+1}) \mapsto (\mathbf{p}^\star_i, \mathbf{p}^\star_{i+1})$ `then`:
        dHPC does reverse sweep of integration with TCs $(\mathbf{S}_{i+1}, \mathbf{p}^\star_{i+1}) \Rightarrow$ optimal plan, $\{\mathbf{s}^\star(t), \mathbf{p}^\star(t)\}_t$, initiating near $\mathbf{S}_i$ ▷ an "origin-directed" reverse SWR-sequence
        commit association $(\mathbf{s}^\star(0), \mathbf{s}^\star(T)) \mapsto (\mathbf{p}^\star(0), \mathbf{p}^\star(T))$ to memory in dHPC ▷ existing episodic memory strengthened
      `else` (memory not found):
        dHPC does an iteration of reverse shooting method from $\mathbf{S}_{i+1} \Rightarrow$ optimal plan, $\{\mathbf{s}^\star(t), \mathbf{p}^\star(t)\}_t$; may not initiate near $\mathbf{S}_i$ ▷ an exploratory reverse SWR-sequence
        commit association $(\mathbf{s}^\star(0), \mathbf{s}^\star(T)) \mapsto (\mathbf{p}^\star(0), \mathbf{p}^\star(T))$ to memory in dHPC ▷ new episodic memory formed
      `if` $\mathbf{s}^\star(0)$ is far from $\mathbf{S}_i$ `then`:
        `break` ▷ SWR train has derailed; terminate trial

Over the course of multiple episodes of awake experience and sleep, these algorithms create a library of optimal plans, leading the system's θ-sequences, which are initially exploratory, to become goal-oriented; a transition that enables the animal to become proficient at navigating the environment.

Further comments are in order regarding the flagging of origins and goals. First, it should be acknowledged that the flagging process has not been included in Algorithm 3; this process would have to be specified, before the algorithm could be considered "complete". Regarding this process, we can say on computational grounds that it should have the following properties. In order to enable the system to focus its resources on computing plans that will be most rewarding, or most relevant to upcoming behavior:

F1. Flagging should come by degrees, which we will call the *priority* of the flagged location; and the selection of $\mathbf{S}_0$ and $\mathbf{S}_g$ in lines 4 and 17 of Algorithm 4 should be biased by priority.

F2. An increase (resp. decrease) in the frequency with which the animal expects to visit a rewarded location in the future, as well as an increase (resp. decrease) in the magnitude of the reward, should increase (resp. decrease) the priority with which the location is flagged as a goal.

F3. Similarly, an increase (resp. decrease) in the frequency with which the animal expects to depart from an unpleasant location in the future, as well as an increase (resp. decrease) in the magnitude of the negative stimulus, should increase (resp. decrease) the priority with which the location is flagged as an origin.

The previous point may be counterintuitive. Imagine a rodent that will sneak out into the open—where it is an easy target for flying predators—in order to reach a source of food. At the point that it has taken the food and is ready to flee, it finds itself departing from a dangerous situation, with its life on the line if it fumbles around; it best have worked out all possible escape plans ahead of time.

F4. Priority for a flagged origin (resp. goal) should decay slightly after every time that location is selected in line 4 (resp. line 17) of Algorithm 4; since fewer optimal plans efferent from (resp. afferent to) that location remain to be computed.

F5. A sudden change to movement affordances in a region (e.g. by opening or closing of a shortcut), as well as a sudden change (increase or decrease) to the cost-rate function $L(\mathbf{s}, \mathbf{a})$ in a region (e.g. by fear conditioning, or the addition of an overhead roof that provides safety from predators), should boost the priorities of all flagged locations (origins and goals) in the vicinity of the region.

The previous point is because such a change in contingencies causes the optimal plans to change, so they need to be re-computed. Finally,

F6. During the awake offline mode, in anticipation of goal-directed navigation, priority should be momentarily boosted for flagged locations (origins and goals) likely to be involved in upcoming plans.

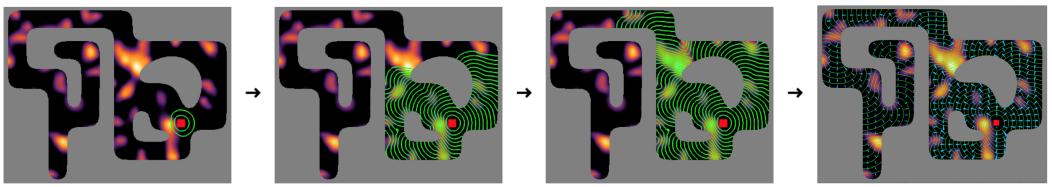
Notice that the selection of flagged locations in lines 4 and 17 of Algorithm 4 influences the SWR-sequences produced by that algorithm. Hence, properties **F1**–**F6** of the flagging process make predictions regarding the frequencies with which forward and reverse SWR-sequences should occur at particular locations, and for the change in such frequencies under various types of manipulations. We will discuss these predictions in Section II.4.3.

Box 3 | Pontryagin vs. HJB: complementary approaches to planning

The brute-force approach to solving our optimal control problem (main text, equations (1)) would be to directly enumerate all valid paths—after suitable discretization of state, time and action—and evaluate each to determine the most efficient one. But the algorithmic complexity of this approach is $\mathcal{O}\left(|\mathcal{A}|^{N_{\text{steps}}}\right)$, exponential in the depth of the search. (Here $|\mathcal{A}|$ is the cardinality of the discretized action space, $|\mathcal{S}|$ that of the discretized state space, and $N_{\text{steps}} \triangleq T_{\text{max}}/\Delta t_{\text{step}}$ is the depth of the search; where T_{max} is the upper limit of the range over which T is allowed to vary, and Δt_{step} is the step-size of discretized time.) Unsurprisingly, this makes brute-force search utterly intractable in practice.

Optimal control theory offers two general approaches for rendering the search tractable, known as the *Hamilton-Jacobi-Bellman* (HJB) *equation* and *Pontryagin's maximum principle* (aka Pontryagin's *minimum* principle, depending on sign conventions) [46]. In the HJB approach, the search is organized by breadth-first, starting at $t = T_{\text{max}}$, and is carried backward recursively in a very efficient way (Bellman's principle of optimality). This leads to a partial differential equation, subject to terminal boundary conditions. The unknown in this equation is the *cost-to-go function* (or its negative, the *optimal value function*). Standard numerical algorithms for solving this terminal boundary value problem have complexity linear in the search depth, $\mathcal{O}\left(|\mathcal{S}| \cdot |\mathcal{A}| \cdot N_{\text{steps}}\right)$, down from exponential. But notice that the proportionality factor here can be quite large, since $|\mathcal{S}| \sim \mathcal{O}(e^n)$ and $|\mathcal{A}| \sim \mathcal{O}(e^m)$ both suffer from the curse of dimensionality.

a



Panel **a** illustrates the features of such an algorithm, using as example the problem from Figure 5 of the main text, and designating as terminal state the particular location marked (red square). Green contours denote level-sets of the cost-to-go function. The far-right panel illustrates the relationship between the underlying "contextual fear" function $\mu(\mathbf{s})$ (heat map), the fully-computed cost-to-go function (contours), and the resulting *optimal policy*, $\mathbf{a}^\star(\mathbf{s})$ (arrows, scaled proportional to $\|\mathbf{a}^\star\|$.) As can be seen, the algorithm appears as value propagating backward through time from the terminal state, and fanning out across space as it propagates.

The HJB equation is closely related to the standard theory of reinforcement learning [212]. The latter provides the theoretical underpinning for an influential paradigm of the procedural (aka habitual) decision-making system [212–215]. It is known to lead to algorithms that are fast-to-act once learning has completed, but which tend to be slow to learn and slow to adapt to changing contingencies [216]. These characteristics are very different from those of the deliberative system; which is flexible in the face of changing contingencies, but does not produce reflex-fast decisions [216, 217]; which relies on depth-first (aka serial) search, not breadth-first [58, 216]; and whose offline computations seem (if SWR-sequences can be interpreted as such) to organize the search not always backward through time, but sometimes backward and sometimes forward. In view of these differences, the HJB approach does not seem to us to provide the best platform for models of the deliberative system.

We turn now to Pontryagin's maximum principle [46]. This approach requires the introduction of a dummy variable, $\mathbf{p}(t) \in \mathbb{R}^n$, the *co-state*, which is time-dependent and of the same dimensionality, n, as the state $\mathbf{s}$; it plays an auxiliary role in the calculations, as we will now see. Define a function, $H(\mathbf{s}, \mathbf{p}; \mathbf{a})$, the *control Hamiltonian*, as

$$H(\mathbf{s}, \mathbf{p}; \mathbf{a}) \triangleq \mathbf{p} \cdot \mathbf{f}(\mathbf{s}, \mathbf{a}) - L(\mathbf{s}, \mathbf{a}). \tag{i}$$

In this approach, calculus of variations is used to directly derive a system of equations which the optimal plan must satisfy, thus sidestepping the search process altogether. (This is analogous to the theorem of ordinary calculus: $[x^\star = \operatorname{argmin}_x g(x)] \Rightarrow [g'(x^\star) = 0]$, which enables replacing a search by an equation.) In this way the minimization problem ends up converted[a] into the

following system of coupled non-linear ordinary differential equations (ODEs):

$$\dot{\mathbf{s}}^\star(t) = \mathbf{f}(\mathbf{s}^\star(t), \mathbf{a}^\star(t)), \tag{iia}$$

$$\dot{\mathbf{p}}^\star(t) = -\frac{\partial H}{\partial \mathbf{s}}(\mathbf{s}^\star(t), \mathbf{p}^\star(t); \mathbf{a}^\star(t)), \tag{iib}$$

$$\mathbf{a}^\star(t) = \underset{\mathbf{a}}{\operatorname{argmax}}\, H(\mathbf{s}^\star(t), \mathbf{p}^\star(t); \mathbf{a}). \tag{iic}$$

These are a system of $2n$ first-order ODEs and m algebraic equations. The unknowns are the $2n+m$ components of the vectors $\mathbf{s}^\star(t), \mathbf{p}^\star(t), \mathbf{a}^\star(t)$, which constitute the optimal plan. The equations must be satisfied for all $t \in [0, T]$. To determine a unique solution, $2n$ boundary conditions are required, which for us are

$$\mathbf{s}^\star(0) = \mathbf{S}_i \quad \text{and} \quad \mathbf{s}^\star(T) = \mathbf{S}_{i+1}. \tag{iid}$$

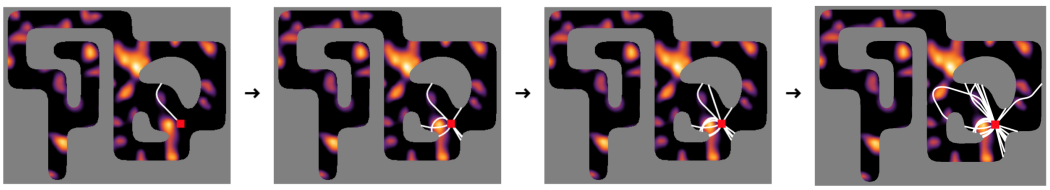
Finally, the duration of the plan, T, is implicitly determined by the equation[b]

$$H(\mathbf{s}^\star(t), \mathbf{p}^\star(t); \mathbf{a}^\star(t)) = U \qquad \forall t \in [0, T]. \tag{iie}$$

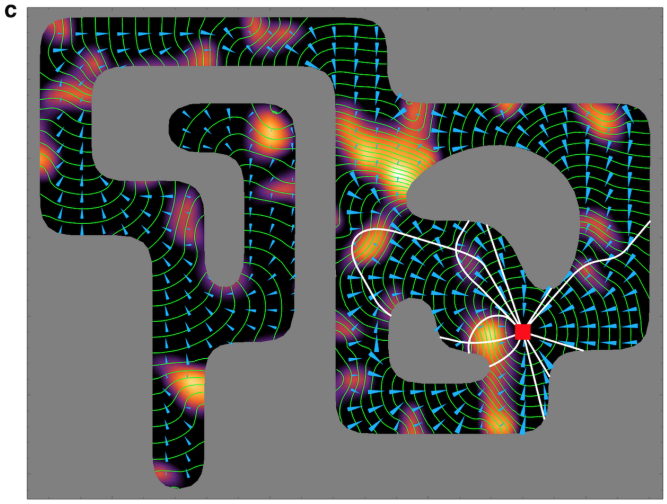
Importantly, notice that the boundary conditions (iid) are split between the two endpoints of integration, making this a two-point boundary value problem (TPBVP). If it were not for this—that is, if the $2n$ boundary conditions could all be given at the initial time $t = 0$—then equations (ii) could easily be solved by a single forward sweep of numerical integration, with a modest algorithmic complexity $\mathcal{O}(|\mathcal{A}| \cdot N_{\text{steps}})$. Instead, TPBVPs must be addressed by some version of the *shooting method*, in which one "shoots" out trajectories in different directions from one boundary until one finds the trajectory that "hits" the other boundary condition [46]. Two straightforward versions of this are the *forward* shooting method, in which one shoots out trajectories "forward through time" from the initial boundary, and the *reverse* shooting method, in which one shoots out trajectories "backward through time" from the terminal boundary.

For an agent who finds themselves needing to navigate between multiple starting and goal locations on a map, a sensible strategy is to implement a "blindfolded" forward shooting method from the starting locations, and a "blindfolded" reverse shooting method from the goal locations. By "blindfolded" I mean that there is no attempt to hit any particular condition at the other boundary. This way, far from having to discard "missed shots", the result is a collection of optimal plans leading from the starting locations out to many places on the map, and from many places on the map in to the goal locations. Panel **b** illustrates an instance of such a blindfolded reverse shooting method, as it builds up optimal plans afferent to one particular goal location (red square).

b



Panel **c** superposes the solutions from Panels **a** and **b**, illustrating that the answers provided by the two approaches are, of course, related.[c] However, it is evident that the two methods organize the computation very differently. In particular, notice that the output of Pontryagin's approach is more modest than that of HJB: the former yields only the optimal plan between a particular start and goal; while the latter yields the optimal policy everywhere in space, afferent to a particular goal. This difference is reflected in the much lower algorithmic complexity of Pontryagin than of HJB. Notice that this makes Pontryagin's approach intrinsically more flexible under changing contingencies than HJB. If a new shortcut opens up, or if a new source of reward appears at some location, it is easier in Pontryagin's approach to discard only the affected trajectories, and begin to recompute. On the other hand, notice also the qualitative similarities between the forward and reverse shooting methods described here, and the coherent paths observed in forward and reverse SWR-sequences in HPC (Section II.2.1 of the main text). This, then, is our main takeaway of this Box: *hippocampal computations appear qualitatively to be consistent with Pontryagin's approach to planning*.



[a]To be precise: Pontryagin's maximum principle provides only a necessary, not a sufficient, condition for global optimality.

[b]One can show that H is constant along any trajectory satisfying (iia–iic). The role of (iie) is simply to determine the value of said constant; which indirectly determines the "time-of-flight", T.

[c]Discrepancies between the two types of solution can be observed in Panel **c** at distances far from the goal. This is related to the notion of cut-loci in Riemannian geometry [218], and has to do with the presence of singularities (e.g. fold and simple cusp catastrophes) in the cost-to-go function [219]. (For reasons of numerical stability, the cost-to-go function plotted is the *viscosity* solution to the HJB equation, which tends to smooth out said singularities [220]; this is why they are not apparent in the contours.) In this regard, it is worth keeping in mind the remark from footnote *a*.

II.3.4 On the mathematics of consolidation

According to our model, an individual episodic memory in HPC is an association $(\mathbf{s}_1, \mathbf{s}_2) \mapsto (\mathbf{p}_1^\star, \mathbf{p}_2^\star)$, between a pair of initial and terminal states and the pair of initial and terminal co-states of the corresponding optimal plan. Notice that the exhaustive collection of all such memories, corresponding to a particular context $\mathcal{S}$, defines a mapping $M : \mathcal{S}^2 \to \mathbb{R}^{2n}$; namely $M(\mathbf{s}_1, \mathbf{s}_2) \triangleq (\mathbf{p}_1^\star, \mathbf{p}_2^\star)$. It happens that M can be greatly compressed, because it is the gradient of a scalar "generating function", $G : \mathcal{S}^2 \to \mathbb{R}$, in the sense that

$$\mathbf{p}_1^\star = -\frac{\partial G(\mathbf{s}_1, \mathbf{s}_2)}{\partial \mathbf{s}_1}, \qquad \mathbf{p}_2^\star = +\frac{\partial G(\mathbf{s}_1, \mathbf{s}_2)}{\partial \mathbf{s}_2}. \tag{4}$$

This function G is just the cost-to-go function (the negative of the optimal value function) appearing in the HJB equation, but viewed now as a function of both initial and terminal states:

$$G(\mathbf{s}_1, \mathbf{s}_2) \triangleq \min_{\{\text{all plans } \mathbf{s}_1 \to \mathbf{s}_2\}} C[\text{plan}], \tag{5}$$

where C is the cost functional from (1b).

This simple mathematical observation may provide the basis for a deeper understanding of the process of consolidation, of hippocampal episodic memory traces into long-term memory during sleep [47].[3] Namely, it suggests that consolidation may be related to the transcription, and simultaneous compression and knitting together, of a collection of associations in HPC of the form $(\mathbf{s}_1, \mathbf{s}_2) \mapsto (\mathbf{p}_1^\star, \mathbf{p}_2^\star)$, into the corresponding cost-to-go

[3]It should be acknowledged that the consolidation process has not been considered in our models above. In particular, the formation and strengthening of episodic memories in Algorithm 4 should not be confused for consolidation.

function, $G(\mathbf{s}_1, \mathbf{s}_2)$, elsewhere in the brain. This rationale is synergistic with, but distinct from, the traditional rationale in terms of learning rates and catastrophic interference [221]. It suggests a powerful way for the deliberative decision-making system to inform and shape the procedural system during sleep, which is consistent with observed place-reward information processing during sleep [87].

II.4 Discussion

II.4.1 Summary

We have proposed an algorithmic model of locale navigation by the deliberative decision-making system. In our model, mPFC and HPC together carry out a hierarchical planning scheme meant to produce flexible decisions in the face of changing contingencies. mPFC leads by computing a high-level plan appropriate to the current context and internal motivations, and then instructs dHPC to "fill in the details" of particular portions of its plan as needed. We've seen how the computational task of dHPC reduces to integrating a system of ordinary differential equations (Pontryagin's maximum principle), and how the split boundary conditions that come with these equations make this a challenging two-point boundary value problem (TPBVP); too difficult to solve all at once at query time. Instead, the system is forced to rely on precomputing the solutions during periods of quiet wakefulness and sleep. In the model, forward and reverse SWR-sequences correspond to times when dHPC is implementing the forward and reverse shooting methods, respectively, in order to solve this TPBVP. The model explains that, after each SWR-sequence, dHPC commits to memory a small "seed" of the solution it just found—which gives a concrete mathematical form to episodic memories. We've explained how this is a memory-efficient way to later allow dHPC to quickly reconstruct the optimal plan at query time, whenever needed, in the form of a θ-sequence. Over the course of multiple episodes of awake experience and sleep, this algorithm builds up a library of optimal plans, leading the system's θ-sequences, which were initially exploratory, to become goal-oriented; a transition that enables the animal to proficiently perform locale navigation in this environment.

Along the way, the model provided functional interpretations for the theta rhythm, the slow oscillation and spindle oscillations; and suggested novel insights into memory consolidation during sleep.

II.4.2 Relationship to previous models

As already mentioned, the three models developed here (together, "our model") draw several of their elements from the models by Eichenbaum [57], Penagos *et al.* [56] and Redish [55]. Namely: from Eichenbaum [57] we have adopted the idea for how, during the awake theta state, thalamus mediates the communication between HPC and PFC; specifically, with vHPC informing PFC of task context, and PFC gating the extraction of relevant information from dHPC to guide behavior. From Penagos *et al.* [56] we have adopted the suggestion that, during sleep, the nested slow oscillation and spindles in PFC, and SWRs in HPC, serve as a means by which complex computations are decomposed into elementary operations across brain regions. And from Redish [55] we have adopted the idea that θ-sequences embody the exploration of future plans, as the basic computation underlying deliberation. For each of the above, related models have been proposed by other authors as well. Our model draws further inspiration from several other works in the literature. For instance: the idea is a classical one that hierarchical planning is key for producing flexible behavior [222]; and it has been argued that anterior PFC [223], particularly anterior mPFC [174], is well-poised to operate at the top of such a planning hierarchy. The general idea that hippocampal sequences are involved in planning has been

explored by several authors [92, 95, 100–106]. And a model by Wang *et al.* [54] is similar in spirit, but not in detail, to our system-level model. At the same time, our model puts forth several novel ideas: that HPC exploits Pontryagin's maximum principle to solve its planning problem—rather than estimating optimal value functions—and the precise manner in which cortical-hippocampal synchrony, and hippocampal sequences, map onto the elements of this algorithm.

It is instructive to compare our model with alternatives which also ascribe to HPC an active role in decision-making. We consider here two such alternatives.

The successor representation

In one influential model, hippocampal place cells are proposed to encode a *successor representation* (SR) of the value function [61, 62]. This model occupies a middle ground between model-based and model-free reinforcement learning; producing decision-making that is flexible under reward revaluation, without sacrificing speed at query time.

Non-overlap. There are several points on which the scope of this model does not overlap with that of ours. For instance, the SR makes a concrete proposal about the "basis" in terms of which individual place cells encode space. As a result, the SR predicts, correctly, that place fields should be shaped by the connectivity of space (as opposed to, e.g., Euclidean distance) [62]. The SR also offers an interesting account of grid cells in entorhinal cortex, as the "eigenvectors of the SR matrix" [62]. We have formulated our model at a higher level of abstraction which is agnostic about the neuronal basis, so our model is silent on these points. In the other direction, the SR has little to say regarding hippocampal sequences, the theta rhythm (including phase relationships of cell firing), interactions between HPC and PFC, or the processing of information during sleep. (But see [63, 99] for extensions of the SR that incorporate roles for SWR-sequences.) Where the scopes of the two models overlap, we find points of agreement and of disagreement.

Agreement. Both models correctly account for an experience-dependent backward expansion of place fields over the course of learning [224], although their explanations differ: in the SR it is a consequence of the fields themselves coming to reflect the animal's policy [62], while in our model it is a miss-attribution of late-phase spikes to the fields; late-phase spikes come to represent the future by virtue of θ-sequences becoming goal-oriented. The two models also succeed in producing decision-making that is flexible under reward revaluations, although it is likely that there are differences to be explored in the degree of this flexibility.

Disagreement. A point of disagreement is the flexibility under changes to movement affordances, such as the opening or closing of shortcuts. Our model is designed to produce flexible decision-making in this situation too, while the SR does not. (But see [99] for extensions of the SR that display such flexibility, at least in simple non-spatial tasks.) Another point of disagreement is the speed of decision-making at query time. In agreement with observations [55], our model produces a sleep-enabled transition, from slow decisions displaying vicarious trial and error early during learning, to streamlined decisions after repeated experience and sleep; while in the SR decisions are always reflex-fast [99]. A third point of disagreement concerns the intrinsic dimensionality, d, of the hippocampal cognitive manifold.[4] The SR predicts $d = n$, the dimensionality of the state space $\mathcal{S}$. In

[4]For a population of N place cells, the population activity at any time can be regarded as a vector $\mathbf{v}(t) \in \mathbb{R}^N$. As the animal explores the state-space $\mathcal{S}$, $\mathbf{v}(t)$ traces out some manifold $\mathcal{S}'$ embedded in $\mathbb{R}^N$. $\mathcal{S}'$ is called the *cognitive manifold* represented by the population of neurons; $d \triangleq \dim \mathcal{S}'$ is its *intrinsic dimensionality*—not to be confused with the dimensionality N of the embedding space.

contrast, as we will discuss in Section II.4.3, our model makes the strong prediction that the dimension of the hippocampal manifold should be $d = 2n$, twice the dimensionality of the state space. Testing this prediction is the subject of ongoing work by the author. The analysis is complicated in $n \geq 2$ due to animal's tendency to develop stereotyped behaviors, which keeps them from exploring all possible trajectories through "high"-dimensional state spaces. For rats running laps on a one-dimensional track ($n = 1$), our analysis (unpublished, not presented here) yields $d = 2$ across datasets from various animals and laboratories, in agreement with our model. This also matches a report of $d = 2$ in a rat performing a one-dimensional ($n = 1$) sound-frequency task [225].

An interesting point of disagreement concerns non-Markovian tasks; in which correct choices are not simply a function of the present state, but depend on past states and must be informed by memory. A simple example of such a task is the alternating T-maze [226]. (A trial on this maze is a run through the stem of the "T", up to the choice point at the junction, and out to either end at the top of the "T". Reward is given only for a left turn after having taken a right turn on the previous trial, and vice versa; so that above-chance performance requires the animal to remember their previous choice.) I claim that such tasks present a two-horned dilemma for value-based models of planning in HPC—to be called the *extravagant* and *austere* horns—as we will now see. The extravagant horn of the dilemma will impale any model which would have HPC cash the value function directly (unlike the SR). For such a model, solving a non-Markovian task would require positing a separate map and value function for each decision-context (the "multiple-map hypothesis" [227, 228]); e.g., in the alternating T-maze, alternating runs would correspond to two different hippocampal maps, each with its own value function. For simple non-Markovian tasks this approach seems computationally appealing; and the remapping problem which it introduces may explain the "trajectory coding" observed in the HPC of trained animals on runs through different routes of T-like mazes [229, 230]. (*Trajectory coding*, aka *trajectory-dependent activity* or *splitter cells*, refers to changes in the firing properties of place cells on different runs through the same place; these changes have been found to depend on the past and future of the animal's trajectory [202, 229, 231–236].) However, an obvious scaling problem emerges as task-complexity grows to ethological levels. To illustrate this, consider the task of Pfeiffer & Foster [105]: a large open arena with 36 clearly separated locations at which reward can be delivered. Trials of exploration and goal-directed navigation are interleaved by baiting a random location on odd trials and a predictable location ("home") on even trials. The home location is chosen at random at the beginning of each day, and remains the same throughout the day. To solve the goal-directed component of this task, value-cashing models would be forced to posit that HPC has 36 separate maps and value functions of this same space; one for each possible home location. It seems unlikely that HPC operates in such an extravagant manner—the extravagant horn of the dilemma. The SR model avoids the extravagant horn by having HPC cash not the value function directly, but the SR matrix, $\hat{M}$, from which the value function can be quickly computed after reward revaluations: $\vec{V}_{\text{new}} = \hat{M}\vec{R}_{\text{new}}$ [62]. Hence, the SR model can solve the goal-directed component of Pfeiffer & Foster [105] using a single hippocampal map (which encodes the matrix $\hat{M}$); it requires only that the home location for the day (encoded in $\vec{R}_{\text{new}}$) be held in memory throughout the day. But now the SR is impaled by the austere horn of the dilemma: since it predicts no trajectory coding on T-like mazes, in contradiction with empirical observations [229, 231–236]. We will discuss in Section II.4.3 how our model avoids both horns of the dilemma: solving non-Markovian tasks by using a single hippocampal map, while providing a framework for trajectory coding in HPC.

The model by Mattar & Daw [63]

A recent model by Mattar & Daw [63], which has received considerable attention, offers another alternative. Their thesis is that planning in HPC, as well as memory processing in HPC during sleep, can both be reduced to model-based learning of a cashed value function; i.e. to reinforcement learning (RL) based on simulated experience, as in the *Dyna algorithm* [212]. They propose a novel prioritization schedule, which they claim to be normative, for the sampling of state-action pairs during simulations. (Samples $(\mathbf{s}, \mathbf{a})$ are used in combination with a forward model, to simulate a transition $(\mathbf{s}, \mathbf{a}) \mapsto (\mathbf{s}', r)$ and perform a one-step Bellman backup, in order to improve the estimate of the value function at $\mathbf{s}$.) Under different circumstances, their schedule produces long forward or backward sweeps of backups, which they offer as a computationally-motivated account of forward and reverse SWR-sequences, respectively. Their proposal is compelling by virtue of the wide range of experimental phenomena that they are able to qualitatively account for [63]. We will refer to this model as *Dyna+MD*.

Computational issues with Dyna+MD. Before delving into a comparison between models, I must say that I have a number of reservations about Dyna+MD, on computational grounds. One reservation is that computing their schedule requires that the agent knows the effect of a backup on its policy prior to deciding whether to perform it, which begs the question. Relatedly, their algorithm, as I understand it, requires that policy evaluation be re-computed (for all states) after every single one-step backup, which is out of the question in terms of computational resources. A third concern is that their updating schedule is only normative in the "myopic" sense (their word) that considers each update in isolation; not in the important sense that would consider the joint effect of multiple updates taken together. In connection to this, notice that a forward chain of k one-step backups is an inefficient way to propagate value information. (Like trying to sweep a floor by doing leftward broom sweeps, while taking steps to the right.) The authors ameliorate this issue by introducing a special rule, which converts the one-step backups in such a forward chain into k-step backups (each of which reaches the end of the chain), but this idea only works for special forward sweeps which happen to replicate the path dictated by the greedy policy at every step; other forward sweeps are left with the problem. In this regard, it is unsatisfactory that the authors only show "performance comparison"[5] with basic Dyna (which makes no effort to schedule samples), not with the stronger Dyna with *prioritized sweeping* (Dyna+PS), which is the standard in RL [212]. Dyna+PS is straightforward to implement efficiently; and while it makes no claim to normativity, its heuristic motivation specifically considers the joint effect of multiple backups taken together; making it the natural null model against which to test their algorithm. These objections weaken the motivation behind their account of SWR-sequences: if HPC could perform Dyna better with other priority schedules, for which there are more efficient algorithms, but which would not reproduce hippocampal phenomenology, then the computational argument works against their model, and empirical evidence is under double burden to support it.

Non-overlap. Setting these worries aside in the hope that they will be addressed in future iterations of the model, we proceed to compare this model to ours. Dyna+MD has a synergy with the SR model [63], and in this way one might say that the predictions of the SR count as predictions of Dyna+MD. Leaving aside these predictions, which we have already discussed, the scope of Dyna+MD is, for the most part, contained within the scope of our model. That is: our model has something to say about most predictions

[5] Since Dyna+MD must be given unrealistic computing powers to overcome the above limitations, even their comparison with basic Dyna needs to be interpreted carefully.

of Dyna+MD; while, in the other direction, our model's predictions regarding the theta rhythm, θ-sequences, interactions between HPC and PFC, and memory consolidation, fall outside the scope of Dyna+MD. Where the scopes of the two models overlap, we find points of agreement and of disagreement.

Agreement. Both models ascribe crucial roles to SWR-sequences in spatial learning. Therefore, they both correctly predict that spatial learning should be correlated with SWR activity [237]; that suppression of plasticity during post-run-sleep [237], as well as disruption of awake [79] or sleep [88, 89, 238] SWRs, should impair spatial learning; and that enhancement of sequence-related SWR activity should improve spatial learning [94]. Furthermore, our prioritization of flagged origins and goals, discussed in Section II.3.3 (points **F1**–**F6**), is meant to solve a similar resource-allocation problem as the priority schedule of Dyna+MD (except at one level above in the planning hierarchy). As a result, both models make a number of similar predictions regarding the frequencies of forward and reverse SWR-sequences at specific times and locations, which are in qualitative agreement with a broad range of observations. These predictions are discussed in Mattar & Daw [63] for one model, and in our Section II.4.3 for the other. They include: (i) a bias for forward over reverse SWR-sequences at the start of a task [76], and (ii) a bias for reverse over forward SWR-sequences after finding reward [76]. (iii) Awake [97] and post-run-sleep [238] SWR-sequence activity increasing with the amount of reward found on the maze, even if the reward was only perceived, without being consumed [239]. Specifically: with reverse SWR-sequence activity at reward locations, but not forward SWR-sequence activity, being monotonically modulated by reward magnitude (increasing as well as decreasing) [98]. (iv) A bias of awake forward SWR-sequences for starting near the animal's location [69, 240], and being predictive of upcoming behavior during goal-directed navigation [105, 241]. (v) A bias of SWR-sequences for specific locations that have been frequently visited [242]. And (vi) novel experiences increasing the incidence of SWRs and associated sequences, both during and after experience, followed by decreasing incidence with increasing familiarity [72, 243–245].

Disagreement. As mentioned earlier, our model produces a sleep-enabled transition, from slow decisions displaying vicarious trial and error (VTE) early during learning, to streamlined decisions after repeated experience and sleep. Presumably Dyna+MD produces a similar course of learning; except that, at query time, decisions must either be reflex-fast or else HPC must go into LIA state to use SWR-sequences to plan. Thus, Dyna+MD in its current state makes no allowance for VTE; which is characterized by HPC remaining in theta state while the animal exhibits behavior of pausing to look in different directions before making its decision [55]. A second point of disagreement concerns the dimensionality, d, of the hippocampal cognitive manifold. As mentioned above, our model predicts $d = 2n$; twice the dimensionality of the state space $\mathcal{S}$. (See Section II.4.3.) Meanwhile Dyna+MD, like the SR, predicts $d = n$. Our prediction matches a report of $d = 2$ in a rat performing a one-dimensional ($n = 1$) sound-frequency task [225]; it also matches the results of an ongoing analysis of one-dimensional spatial tasks by the present author (unpublished, not presented here). Cases with $n \geq 2$ have yet to be carefully tested. A third point of disagreement concerns non-Markovian tasks, discussed previously in connection to the SR. We will discuss in Section II.4.3 how our model solves such tasks. Meanwhile, Dyna+MD is impaled by one or the other (austere or extravagant) horn of the dilemma discussed previously; depending on whether it is used to learn the SR or to learn the value function directly [63]. A potential fourth point of disagreement is whether SWR-sequences represent smooth (i.e. differentiable) trajectories through space, or are more akin to drift-diffusion processes (aka Brownian motion). Dyna+MD predicts sequences

that sometimes resemble the (smooth) trajectories of the animal [63], but it is unclear to me wether smoothness is a categorical prediction of the model. As we will discuss in Section II.4.3, our model categorically predicts smooth trajectories; which is in agreement with a recent report on awake SWR-sequences [241], but in tension with an earlier report on sleep SWR-sequences [246]. (See discussion in Section II.4.4.)

II.4.3 Predictions and further interpretation

In this Section we discuss various additional predictions and matters of interpretation of our model.

Propensities of SWR-sequences

When comparing our model to Dyna+MD in Section II.4.2, we mentioned a number of predictions regarding the propensities of forward and reverse SWR-sequences under various conditions. Here we briefly explain where each of these predictions comes from. As mentioned in Section II.3.3, the selection of flagged locations in lines 4 and 17 of Algorithm 4 influences the SWR-sequences produced by that algorithm. In this way, the prioritization properties **F1**–**F6** of the flagging process (Section II.3.3) influence the propensities of SWR-sequences, giving us the predictions in question. We go over this list of predictions in the order they were mentioned in Section II.4.2.

(i) A bias for forward over reverse SWR-sequences at the start of a task [76].

This follows from **F6**, which recommends temporarily boosting the priority of the animal's current location as a flagged origin, in anticipation of upcoming goal-directed navigation, since this will certainly be a point of origin for upcoming plans.

(ii) A bias for reverse over forward SWR-sequences after finding reward [76].

This follows from **F2**, which recommends increasing the priority of the rewarded location as a flagged goal, since finding a reward at the location is a good heuristic predictor that there will be reward there again in the future.

(iii) Awake [97] and post-run-sleep [238] SWR-sequence activity increasing with the amount of reward found on the maze, even if the reward was only perceived, without being consumed [239]. Specifically: with reverse SWR-sequence activity at reward locations, but not forward SWR-sequence activity, being monotonically modulated by reward magnitude (increasing as well as decreasing) [98].

This follows directly from **F2**. However, in the case of decreasing reward magnitude there is a counteracting effect from **F5**, which recommends increasing the priority of the rewarded location as a flagged goal. Hence, we can predict that the increase in reverse SWR-sequence activity resulting form an increase in reward magnitude should be more pronounced than the decrease in the same resulting form a decrease in reward magnitude. This further prediction is also borne out experimentally [98].

(iv) A bias of awake forward SWR-sequences for starting near the animal's location [69, 240], and being predictive of upcoming behavior during goal-directed navigation [105, 241].

This follows from **F6**, the same as in point (i) above.

(v) A bias of SWR-sequences for specific locations that have been frequently visited [242].

This follows from **F2** and **F3**, since the frequency with which a flagged location has been visited in the past is a good heuristic predictor of the frequency with which it will be visited in the future.

(vi) Novel experiences increasing the incidence of SWRs and associated sequences, both during and after experience, followed by decreasing incidence with increasing familiarity [72, 243–245].

This follows from **F5**, which recommends increasing the priorities of flagged locations in the vicinity of places where contingencies have changed; and from **F4**, which recommends gradually decreasing the priority of flagged locations as Algorithm 4 proceeds to compute more of the optimal plans to/from those locations.

We mention two further predictions regarding SWR-sequence propensities, which, to my knowledge, have not yet been tested experimentally.

(vii) Remote SWR-sequences (those not stemming from the location of the animal) should stem from discrete fixed locations on the map.

This follows from Algorithm 4, in which forward and reverse SWR-sequences always stem from the discrete points $\{\mathbf{S}_1, \mathbf{S}_2, \dots\}$ making up the high-level map (see Figure 3). The exception is the present location of the animal ($\mathbf{S}_0$) during the awake offline mode; which, of course, is not constrained to the discrete points making up the high-level map, yet can serve as a point in the high-level plan (see Figure 3).

(viii) For an animal on a novel maze, SWRs should occur predominantly as singlets (SWR trains of length one). SWR trains should gradually become longer (lengths two, three, etc.) over repeated episodes of experience and sleep, simultaneously as the animal becomes proficient at navigating the maze.

This follows from Algorithm 4. On a novel maze the collection of episodic memories (3) starts out empty; so the algorithm produces only exploratory SWR-sequences (lines 12 and 25), which tend not to hit the required boundary condition to continue the train in the multiple shooting method, triggering the `break` clause in lines 14–15 and 27–28. Over the course of learning the algorithm builds up its collection of episodic memories, allowing it to produce directed SWR-sequences more often (lines 9 and 22), which tend to hit the required boundary condition to continue the train.

Role of θ-sequences in deliberation

We mention a couple of predictions regarding the role of θ-sequences in deliberation.

(ix) θ-sequences should transition, from being exploratory early during learning of a novel maze [74], to becoming goal-directed after sufficient experience and sleep [55], hand in hand with the animal becoming proficient at the task [55].

This follows from Algorithms 3 and 4. On a novel maze the collection of episodic memories (3) starts out empty; so Algorithm 3 produces only exploratory θ-sequences (line 7). Over the course of learning Algorithm 4 builds up its collection of episodic memories, allowing Algorithm 3 to produce goal-directed θ-sequences more often (line 5). This enables the system to plan successfully in real time.

(x) Disruption of θ-sequences should disable the animal's ability for locale navigation. This should appear as a decrease in the animal's performance on tasks which can only be solved by locale navigation and not by any of the other strategies for navigation (cf. Section II.2.3). On tasks that can be solved by either locale or route

navigation, the disruption should cause the animal to switch to route navigation; i.e. favoring taxon navigation between subgoals, over the nuanced avoidance of intervening spaces which are exposed or have been fear-conditioned. Because route navigation is computationally easier than locale navigation, this may manifest, counterintuitively, as an improvement in performance as measured in trials-per-minute, or even percentage-correct-trials—even if it is an inferior strategy under ethological conditions.

This follows from the function of θ-sequences in Algorithm 3, as reconstructions at query time of imminently-needed optimal plans. These are low-level plans, used to enable execution of the high-level plan of mPFC. As mentioned in Section II.3.2, taxon navigation can serve as a HPC-independent substitute for these low-level plans,[6] but has the disadvantage of exposing the animal to regions of space that should be avoided under ethological conditions. To my knowledge, prediction (x) has not been carefully tested. (But see Siegle & Wilson [247], and further discussion in Section II.4.4.)

Smoothness of hippocampal sequences

(xi) Hippocampal sequences (both θ- and SWR-) should follow smooth (i.e. differentiability class C^1 almost-everywhere) paths, $\{\mathbf{s}(t)\}_t$; not jagged or Brownian-motion-like.

This follows from the numerical role of hippocampal sequences in our model, as instances when dHPC is integrating the ODEs of Pontryagin's maximum principle ((iia–iic) of Box 3) either forward or backward through time. These ODEs imply that $\dot{\mathbf{s}}(t)$ exists and is an almost-everywhere continuous function of time (and hence $\mathbf{s}(t)$ is continuous). Prediction (xi) is in agreement with a recent analysis of awake SWR-sequences [241], and in tension with an earlier analysis of sleep SWR-sequences [246]. (But see discussion in Section II.4.4.)

Functions, informational contents, and geometries of cognitive manifolds, for the subfields DG, CA3 and CA1

Our model says that the input to dHPC is a pair $(\mathbf{S}_0, \mathbf{S}_1)$ of initial and terminal states (Figure 4c). Accordingly, we might expect this to be the information represented at the input of dHPC, in dentate gyrus (DG):

(xii) The DG field of dHPC should represent the cognitive manifold $\mathcal{S}^2 = \{(\mathbf{S}_0, \mathbf{S}_1)\}$, where $\mathbf{S}_0$ represents the current state of the animal, and $\mathbf{S}_1$ the immediate subgoal of the high-level plan from mPFC.

Admittedly, this prediction is vague regarding which of the two principal cell populations of DG—granule cells or mossy cells [248]—should instantiate the manifold in question.

The representation $(\mathbf{S}_0, \mathbf{S}_1)$ is not suitable for numerical integration of the ODEs ((iia–iic) of Box 3). For that, $(\mathbf{S}_0, \mathbf{S}_1)$ should be transformed to $(\mathbf{S}_0, \mathbf{p}_0^\star)$, where $\mathbf{p}_0^\star$ is the initial co-state of the corresponding optimal plan, as in (3). A natural place within HPC for the integration of the ODEs to be computed is within the recurrent connections of CA3.[7]

[6] Refer to Redish *et al.* [49] for a review of anatomical structures involved in taxon navigation.

[7] Numerical integration of ODEs can naturally be performed by a recurrent neural network (RNN). Let $\vec{\pi}_0$ be a vector of initial probabilities over a set of discretized values for $(\mathbf{s}, \mathbf{p})$. Pontryagin's ODEs, which define a flow over the space $\{(\mathbf{s}, \mathbf{p})\}$, induce a linear dynamics on $\vec{\pi}$; that is, $\dot{\vec{\pi}} = \hat{M}\vec{\pi}$, where $\hat{M}$ is some matrix of coefficients (the transition matrix). Integrating this ODE over a small finite time step $\Delta t = 1$ (in some units):

$$\vec{\pi}_{t+1} = \hat{P}\vec{\pi}_t, \tag{6}$$

where $\hat{P} = \exp\{t\hat{M}\}$ is a Markov matrix. By encoding $\hat{P}$ in the synaptic connections of a linear RNN, the RNN's dynamics will precisely replicate (6), solving Pontryagin's ODEs while elegantly handling uncertainty in the initial conditions.

Accordingly, we might expect that

(xiii) The CA3 field of dHPC should represent the cognitive manifold $\mathcal{S} \times \mathbb{R}^n = \{(\mathbf{s}, \mathbf{p})\}$.[8] During the awake theta state, at the start of the theta cycle, $\mathbf{s} = \mathbf{S}_0$ represents the current state of the animal and $\mathbf{p} = \mathbf{p}_0$ the initial co-state of a candidate optimal plan starting at $\mathbf{S}_0$.

Furthermore (cf. footnote 7),

(xiv) The CA3 field of dHPC should function as a numerical integrator for the ODEs of Pontryagin's maximum principle ((iia–iic) of Box 3). Having received as input from DG a full set of initial conditions for these ODEs, $(\mathbf{S}_0, \mathbf{p}_0)$, CA3 should, over the course of a theta cycle or a SWR, compute an optimal plan $\{(\mathbf{s}(t), \mathbf{p}(t))\}_t$ from those initial conditions; a hippocampal sequence.

This prediction pinpoints CA3 as the originator of hippocampal sequences. Consistent with this prediction, SWRs in CA1 are known to be induced by CA3 activity [249]—suggesting an interpretation in which SWR-sequences originate in CA3 before being communicated to CA1. In agreement with this interpretation, chronic blockade of CA3 to CA1 transmission resulted in a loss of SWR-associated reactivation in CA1 [250]; and a more targeted acute silencing of CA3 revealed its dominant role in CA1 place field responses and ensemble activity [251]. (The latter experiment found no effect of CA3 silencing on CA1 theta phase precession, but it remains to be seen whether this entails intact CA1 θ-sequences—if it did, that would be in tension with our prediction (xiv).)

We have just seen that a likely prediction of our model is that, over the course of each theta cycle or SWR, CA3 computes a trajectory through $\{(\mathbf{s}, \mathbf{p})\}$-space. Recall that the co-state, $\mathbf{p}$, is an auxiliary variable necessary for working with Pontryagin's ODEs. Once these ODEs have been solved, this auxiliary variable has served its purpose and can be discarded. Thus it would make sense for CA3 to communicate only $\{\mathbf{s}(t)\}_t$ downstream to CA1. Now, of the computational task ascribed to dHPC by our model, the only part we are still missing is the validation step at the output (step **f** of Box 2, and lines 14-15 and 27-28 of Algorithm 4), where it is checked whether the produced low-level plan $\{\mathbf{s}(t)\}_t$ indeed terminates near the requisite subgoal of mPFC, $\mathbf{S}_1$. CA1 is well poised to carry out this final validation, since: (i) CA1 is located at the output of HPC; (ii) CA1 receives input from area CA3 (which we have just argued conveys $\{\mathbf{s}(t)\}_t$) as well as from mPFC (through RE, which conveys $\mathbf{S}_1$); and (iii) CA1 projects back to mPFC (both through RE and directly, cf. Box 2), as would be needed to convey the output of the validation back to mPFC (cf. Box 2). Hence, we might expect that

(xv) The CA1 field of dHPC should represent the cognitive manifold $\mathcal{S}^2 = \{(\mathbf{s}(t), \mathbf{S}_1)\}$, where $\mathbf{s}(t)$ is the output from CA3, which sweeps out a sequence through $\mathcal{S}$ once per theta cycle or SWR, and $\mathbf{S}_1$ is the immediate subgoal of the high-level plan from mPFC.

It follows from predictions (xii, xiii, xv), in particular, that for a task with an n-dimensional state-space $\mathcal{S}$,

(xvi) The cognitive manifolds represented in the DG, CA3 and CA1 fields of dHPC should each have intrinsic dimension $d = 2n$; twice that of the task's state-space.

This prediction matches a report of $d = 2$ for the CA1 cognitive manifold in a rat performing a one-dimensional ($n = 1$) sound-frequency task [225]. It also matches the results of an

[8] Strictly speaking, the manifold $\{(\mathbf{s}, \mathbf{p})\}$ is the *cotangent bundle* of $\mathcal{S}$, denoted $T^*\mathcal{S}$. This is always locally, but not always globally, isomorphic to the cartesian product $\mathcal{S} \times \mathbb{R}^n$.

ongoing analysis of one-dimensional spatial tasks by the present author (unpublished, not presented here). CA1 data from a rodent foraging in an open arena ($n = 2$) was analyzed by Low *et al.* [225], who report finding $d = 3$. This does not quite match our prediction of $d = 4$; however, it is possible that this may reflect incomplete sampling of the cognitive manifold in this experiment. Indeed, cases with $n \geq 2$ need to be carefully tested; these are complicated by animal's tendency to develop stereotyped behaviors, which keeps them from exploring all possible trajectories through "high"-dimensional state spaces.

Locus of episodic memory traces

As has just been discussed, a likely prediction of our model is that the transformation between representations $(\mathbf{S}_0, \mathbf{S}_1)$ and $(\mathbf{S}_0, \mathbf{p}_0^\star)$ takes place in the passage from DG to CA3. As has been explained earlier in connection with equation (3), in our model this very transformation is the content of episodic memory. Accordingly, we might expect that

(xvii) The locus of episodic memory traces (3) should be the synaptic connections of the DG's mossy fibers onto CA3 dendrites.

Senzai [248] reviews the hypotheses that DG functions as a pattern separator, while CA3 functions as a pattern-completing auto-associative network, and that the locus of episodic memory traces are the recurrent connections of CA3. Our model's predictions are different from these hypotheses, but we can see how it could look like that if our model's predictions were right: notice that an episodic memory in our model is only the "seed" of an episode; the episode itself gets reconstructed from this seed in CA3.

Solving non-Markovian tasks

The general principle by which our algorithm is able to solve non-Markovian tasks is quite simple. We illustrate it on the alternating T-maze. To solve this task our algorithm requires one hippocampal map (not two); together with mPFC's ability for representing the necessary task rule (alternation), and mPFC's access to adequate working memory. The latter assumptions are both empirically supported [163, 164, 252]. Using these resources, mPFC can use classical planning techniques for its high-level planning (Figure 3), which takes care of the non-Markovian aspect of the task. The contribution of HPC remains the same as in Markovian tasks: to work out the imminently-needed details of the high-level plan. It is easy to see that this method of solution evades the extravagant horn—a bad scaling of resources with number of decision contexts involved—of our extravagant-austere dilemma, laid out in Section II.4.2. For example, to solve the goal-directed component of the task by Pfeiffer & Foster [105], with its 36 distinct possible home locations, our algorithm still requires just one hippocampal map; it is enough that PFC have access to a memory of the home location throughout the day, so that it can perform its high-level planning accordingly.

Trajectory coding in HPC

We have just seen that our model avoids the extravagant horn of our extravagant-austere dilemma, by requiring only one hippocampal map to solve a given non-Markovian task, no matter how many decision contexts may be involved. In this sense our model resembles the successor representation (SR) model. But we saw in Section II.4.2 that the SR gets impaled by the austere horn of the dilemma—an inability to accommodate trajectory coding in HPC.[9] As we will now explain, our model avoids also this horn of the dilemma.

[9] As a reminder, trajectory coding (aka trajectory-dependent activity or splitter cells) refers to changes in the firing properties of place cells on different runs through the same place. These changes have been found to depend on the past and future of the animal's trajectory [202, 229, 231–236].

Consider first the dentate gyrus (DG) field of dHPC. According to our prediction (xii), during the awake theta state, neurons in DG should encode not just the present location of the animal, $\mathbf{S}_0$, but also, conjunctively, the immediate subgoal of the high-level plan from mPFC, $\mathbf{S}_1$. Moreover, since these data simply reflect the input to dHPC, they should be manifest as soon as the map emerges. This means that

(xviii) DG place cells should exhibit trajectory-dependent prospective firing. This tuning property should be manifest as soon as the map emerges; after only a few runs through a novel environment.

(As in prediction (xii), this prediction is admittedly vague regarding which of the two principal cell populations of DG—granule cells or mossy cells [248]—should exhibit this tuning.) *Prospective* (resp. *retrospective*) firing means that the neuron's activity depends on the future (resp. past) animal trajectory. Trajectory coding in DG was indeed reported by Senzai & Buzsáki [253], but as far as I'm aware no studies have been done to distinguish whether this activity is prospective or retrospective, nor how soon this tuning emerges in a novel environment.

We turn now to the CA3 field of dHPC. According to our prediction (xiii), during the awake theta state, at the start of the theta cycle, neurons in CA3 should encode not just the present location of the animal, $\mathbf{S}_0$, but also the initial co-state, $\mathbf{p}_0$, of a candidate optimal plan starting from $\mathbf{S}_0$. Early during learning of a novel environment, before episodic memories (3) have been created, this co-state $\mathbf{p}_0$ will be chosen at random according to line 7 of Algorithm 3, and will bear little relation to the animal's future trajectory. Hence,

(xix) Early during learning of a novel environment, CA3 place cells should exhibit no (or only weak) trajectory-dependent prospective firing.

However, over the course of multiple episodes of awake experience and sleep, as dHPC builds up its collection of episodic memories, this will enable Algorithm 3 to call on line 5 more often. This transition, which causes θ-sequences to become goal-oriented, means that $\mathbf{p}_0$ becomes predictive of the PFC's immediate subgoal $\mathbf{S}_1$. Hence,

(xx) After multiple episodes of awake experience and sleep, hand in hand with θ-sequences becoming goal-oriented, CA3 place cells should develop trajectory-dependent prospective firing.

As far as I'm aware, no longitudinal studies have been reported to determine if, and how soon, trajectory coding in CA3 emerges in a novel environment. However, Ito *et al.* [202] reported weak prospective coding in CA3 (in rats trained until achieving 90% correct trials on the alternating T-maze); which may represent a single snapshot along the transition, from no prospective coding initially (prediction (xix)) to robust prospective coding as θ-sequences become goal-oriented (prediction (xx)).

Finally we turn to the CA1 field of dHPC. According to our prediction (xv), during the awake theta state, at the start of the theta cycle, the coding in CA1 is much like that in DG: neurons in CA1 should encode the present location of the animal, $\mathbf{S}_0$, conjunctively with the immediate subgoal of the high-level plan from mPFC, $\mathbf{S}_1$. And since these data simply reflect the input to dHPC, they should be manifest as soon as the map emerges. Therefore

(xxi) Like DG place cells, CA1 place cells should exhibit trajectory-dependent prospective firing. This tuning property should be manifest as soon as the map emerges; after only a few runs through a novel environment.

This prediction is in agreement with robust trajectory-dependent prospective firing observed in CA1 [202, 233, 236]. It may also explain why CA1 place cells become directionally selective on linear tracks [254, 255].

We can make an additional prediction about *where* in space prospective tuning should occur (for any subfield of HPC). We have seen that such prospective tuning comes about in our model due to HPC encoding not only the present state, $\mathbf{S}_0$, but also, conjunctively, the immediate subgoal, $\mathbf{S}_1$, of the high-level plan from mPFC (or in the case of CA3, the corresponding initial co-state, $\mathbf{p}_0$, which comes to reflect $\mathbf{S}_1$ after the necessary episodic memories (3) have been formed). Since $\mathbf{S}_1$ is the *first* upcoming subgoal of the high-level plan, not any subgoal further ahead:

(xxii) Hippocampal trajectory-dependent prospective firing should never be present further back from the choice point than the typical distance between the points making up the high-level map in mPFC (as in Figure 3); this should be the same as the typical length of a simple SWR-sequence ($\sim$ 50cm in rodents).

This prediction has been tested and is consistent with observations [256].

In this way, our model provides a framework for prospective trajectory coding in HPC. We note, however, that hippocampal place cells are also known to exhibit retrospective trajectory coding [229, 232, 234, 235]. It does not seem that our model provides an explanation for this phenomenon, which might best be addressed by the "multiple-map hypothesis" [227, 228, 256].

II.4.4 Limitations and points of tension

In this section we make note of several limitations in the scope of our model, and of some points of tension with available empirical evidence.

Limitations of scope

As any other model, our model is idealized in several ways and is limited in its scope. For example: (i) as regards locale navigation, we have entirely neglected the key problems of mapping;[10] of model learning;[11] of fear/reward conditioning;[12] and of self-localization.[13] We hope that thinking of our model as embedded in a larger system (cf. Box 2) may suggest connections to existing as well as new hypotheses, for how the deliberative system as a whole solves such problems. (ii) An important aspect of episodic memory is the binding together of multimodal stimuli into a context in the form of a wholistic episode [47]. Our model has offered a mathematical characterization of episodic memories which seems to capture the sequential or "mental time travel" quality of such memories, but we have had little to say about the binding together of multimodal stimuli. This omission relates to the previously mentioned one regarding the problem of mapping. (iii) Our model posits only two levels of hierarchical planning, as in Figure 3, but it may be possible to modify our model to include more levels. Indeed, the grading of spatial scales observed along the long axis of the hippocampal formation [155] may suggest such a modification. (iv) As mentioned at the end of Section II.4.3, our model does not seem to provide an explanation for the observed retrospective trajectory coding in HPC [229, 232, 234, 235]. This phenomenon might best be addressed by the "multiple-map hypothesis" [227, 228, 256].

Consistent with the observations highlighted in Section II.2.1, our model distinguishes two modes of operation for the deliberative system: online and offline. However, it is likely that these modes subdivide further in ways our model does not capture. Studies of sleep indicate three or four distinct stages of non-REM sleep [257]. Our model's offline mode

[10] i.e. learning the state space $\mathcal{S}$.

[11] i.e. learning the action space $\mathcal{A}$ and the transition function $\mathbf{f}$.

[12] i.e. learning the cost-rate function L.

[13] i.e. determining the current context, $\mathcal{S}$, and the current state, $s \in \mathcal{S}$, and handling remaining uncertainty.

(Algorithm 4) may describe the computations involved in only one of those stages (e.g. stage 2), while memory consolidation (Section II.3.4) may take place during a different stage of sleep (e.g. stage 3). Concerning the waking state, Wu *et al.* [240] performed strong fear conditioning at one end of a track in rats running laps on a linear track. Before conditioning, awake SWR-sequences were predictive of the animal's upcoming behavior, consistent with our model (prediction (iv) of Section II.4.3) and with other studies [105, 241]. However, after strong fear conditioning SWR-sequences became predictive of paths that the animal *avoided*. This happened hand in hand with an overall reduction in the animal's running speed. These observations suggest either a discrete transition in the mode of operation of the deliberative system, from goal-directed to fear-aversion—in which case our model would describe only the goal-directed mode—or else a takeover of decision-making by a fear-aversion system, distinct from the deliberative system.

Tensions with empirical evidence

We have treated SWR-sequences as being always perfectly forward or perfectly reverse, but in fact a fraction of SWR-sequences are mixtures—part forward, part reverse [69]—as seen in Figure 1e. Within our model, perhaps this indicates a small error rate in Algorithm 4, which allows jumping between the two `for` loops in lines 6 and 19.

Siegle & Wilson [247] used optogenetic stimulation to inhibit dorsal CA1 at specific phases (either early or late) within each theta cycle, and tested the effect on task performance when either manipulation was done during the context-cueing or memory-retrieval phases of a spatial working-memory task. Surprisingly, they found combinations in which their manipulation *improved* performance (specifically, when CA1 was inhibited at early phases of theta during context cueing, as well as when it was inhibited at late phases of theta during memory retrieval). On first impression this result seems to conflict with any model in which θ-sequences play an active role in decision-making (as in our model). Any such model would predict that disruption of θ-sequences should negatively impact the system's performance. Seemingly adding to this conflict are observations that rats are perfectly capable of solving the alternating T-maze (without delay) after complete lesioning of their HPC [258]. However, it is possible that there is no conflict here. As highlighted in prediction (x) of Section II.4.3, our model implicates θ-sequences, and HPC generally, in *locale* navigation, but there are other HPC-independent strategies for navigation available to the brain (Section II.2.3). In particular, the alternating T-maze and the task employed by Siegle & Wilson [247] can both be solved by taxon navigation. So it is possible that the sustained (or even improved) task performance they report after disruption of HPC is a reflection of the brain switching to taxon navigation.

II.4.5 Future directions

In the near term, I'm interested in the questions of (i) how the hierarchy is extracted; (ii) how to incorporate uncertainty into the model (see Part III); (iii) what is the neuronal-level implementation of our algorithm; (iv) how, in detail, does memory consolidation work (cf. Section II.3.4); and (v) beyond spatial navigation, what can our model say about goal-directed decision-making in more abstract domains? I also look forward to collaborating with my experimental colleagues to test the predictions from Section II.4.3.

In the long term, the highest form of success for our model would have it become established as a paradigm—a conceptual framework—guiding the community in understanding the "computational life" of hippocampus, and its relation to the larger deliberative system. In such a happy scenario, our model would do for the deliberative system what the theory of reinforcement learning has done for the habit system [212–215]. Much work would need to be done in pinning down the model's loose ends; this would require, in particular, the

many experimental labs in the field, with their myriad windows into the deliberative system, to grapple for alignment between our model and their views. Naturally, this would involve experiments testing our model from many angles, including those highlighted in Section II.4.3.

Part III

A conjecture on intertemporal choice

> Perhaps the structure of [quantum] theory denotes the optimal way to reason and make decisions in light of some fundamental situation—a fundamental situation waiting to be ferreted out in a more satisfactory fashion.
>
> C. Fuchs [39]

In this concluding part of the thesis, I attempt to explain how the two main parts (Parts I and II) fit into the context of a broader research agenda. I make no attempt here at establishing results. Instead I present a thread of reasoning involving a number of speculative—but I hope, stimulating—connections. The discussion involves ideas from a diversity of fields: Hamiltonian mechanics, symplectic geometry, optimal control theory, reinforcement learning, probability theory, decision theory, quantum mechanics and cognitive science. For the sake of brevity, I've decided against stopping to define the various concepts and equations invoked (as long as they are well-established in their fields); limiting myself to providing entries to the literatures as appropriate. The punchline of this part will be a conjecture, that the probability calculus of quantum mechanics holds a kind of normative status for a class of decision problems involving intertemporal choice under uncertainty—a class of problems of great importance to artificial intelligence, brain sciences, economics, and, I argue, to physics too.

III.1 Symplectic geometry as the implicit common thread of this thesis

We begin by pointing out *symplectic geometry* as a common element underlying the mathematics in the two main parts of this thesis. On the one hand we have Hamiltonian mechanics (the subject of Part I), which is widely considered to be one of the crown jewels of mathematical physics, and which is a confluence of differential, algebraic and symplectic geometry, Lie algebra and Lie groups [17]. All of this structure follows from little more than Hamilton's principle of stationary action, which states that if an isolated physical system evolves from configuration q_1 to configuration q_2 over a period of time $[t_1, t_2]$, it must do so along a trajectory which makes the action functional stationary (i.e. such that the first variation $\delta S = 0$):

$$\text{make stationary} \qquad S[q] = \int_{t_1}^{t_2} dt\, L(q, \dot{q}; t), \tag{1a}$$

$$\text{between} \qquad q(t_1) = q_1 \text{ and } q(t_2) = q_2. \tag{1b}$$

(Here L is the Lagrangian function for the system, and S the action functional.) On the other hand, optimal control theory (the mathematical underpinning for Part II) is

concerned with solving the optimal control problem [46]:

$$\text{control the system} \qquad \dot{q} = f(q, a), \tag{2a}$$

$$\text{to minimize} \qquad S[a] = \int_{t_1}^{t_2} dt\, L(q, a; t), \tag{2b}$$

$$\text{between} \qquad q(t_1) = q_1 \text{ and } q(t_2) = q_2. \tag{2c}$$

(Here $q(t)$ is the state of the system, $a(t)$ is the control, or action, $f(q, a)$ is the transition function specifying the effect of taking action a in state q, $L(q, a; t)$ is the cost-rate for taking action a in state q at time t, and $S[a]$ is the cost of a given trajectory.) I've used overlapping notations in (1, 2) to emphasize that the two mathematical problems are closely interrelated. Indeed, the reader may be unsurprised to hear that the theory of optimal control reuses (or extends) many of the same concepts and equations as Hamiltonian mechanics. Thus the Lagrangian becomes the cost-rate; the canonical momentum becomes the co-state; the Hamiltonian becomes the control Hamiltonian; Hamilton's equations become Pontryagin's minimum principle; the Hamilton-Jacobi equation becomes the Hamilton-Jacobi-Bellman (HJB) equation; and Hamilton's principal function becomes the cost-to-go function. Naturally, the rich symplectic geometry of Hamiltonian mechanics also underlies the mathematics of optimal control. This prevalence of symplectic geometry in variational problems was stated with authority by Arnold and Givental [259]:

> Whenever the equations of a theory can be gotten out of a variational principle, symplectic geometry clears up and systematizes the relations between the quantities entering into the theory. Symplectic geometry simplifies and makes perceptible the frightening formal apparatus of [...] the calculus of variations in the same way that the ordinary geometry of linear spaces reduces cumbersome coordinate computations to a small number of simple basic principles.

III.2 Exploiting symplectic geometry in reinforcement learning

I want to emphasize the great practical value of exploiting symplectic structure when it is present in a problem. For this, recall some of the many momentous results in mechanics that exploit symplectic structure: the existence of local canonical coordinates on any symplectic manifold (Darboux's theorem) [17]; the Poisson bracket, the algebra of observables, and the relation between symmetries and conservation laws (symplectic Noether's theorem) [17]; the invariant measure on phase space (Liouville's theorem) [17], with its far-reaching consequences in statistical mechanics [21]; the theory of integrable systems, invariant tori, and action-angle coordinates (Liouville-Arnold theorem) [17]; the stability of invariant tori, and the structure of soft chaos, in perturbed integrable systems (KAM theorem) [17]; and symplectic numerical integration schemes, which vastly outperform their non-symplectic counterparts [260, 261]. I'd also like to contrast this list with how little, as of yet, symplectic structure has been recognized and exploited outside of physics and mathematics. Indeed, as far as I'm aware, the only exploit of symplectic geometry in optimal control has been a tepid adoption of symplectic integrators when optimal control needs be exerted over prolonged periods of time [262–267].

With the thought that "symplectic geometry is a powerful resource" on our minds, the problem that I would like to draw attention to is the following.

P: How does one solve a variational problem of the form (1) or (2) in a situation where there is incomplete information available about the boundary conditions, q_1, q_2, the transition function f, and/or the cost-rate function L?

This type of problem is of great interest in artificial intelligence, particularly in reinforcement learning (RL); where an agent that begins with little to no prior knowledge tries to accrue as much reward as possible (i.e. to solve (2)), while simultaneously trying to infer $q(t)$ from noisy sensory data and to learn f, L through experience. The standard approach in RL has been to model the problem as a discrete-space, discrete-time, Markov decision process (MDP), with unknown transition and reward probabilities [212]. That is: state transitions and rewards are described as random variables, so that taking action a in state s leads to state s' and reward r, not deterministically, but at random, according to some objective and initially-unknown probability distribution $p(s', r|s, a)$. The power of that approach, as far as I can tell, comes from the Bellman optimality equation—a version of the HJB equation suited to discrete MDPs—and from the dynamic programming techniques stemming from it [212]. (Among those techniques are staples of the field such as value back-propagation and temporal-difference learning.) However, it seems to me that a great resource is wasted in replacing the deterministic-but-uncertain problem, **P**, with the inherently-stochastic MDP version of the problem. Namely, problem **P**—that which remains of the optimal control problem (2) "modulo" knowing the exact value of q_1, q_2, f, L—still has symplectic structure (we know the optimal trajectories must define *some* symplectic flow on phase space; although we may not know exactly which flow nor where we are in phase space); but that structure is completely absent from the MDP version of the problem. By analogy to the way symplectic integrators are able to vastly outperform conventional integrators by exploiting symplectic structure [260, 261], I expect there must be symplectic algorithms for solving **P** which vastly outperform current RL algorithms.

III.3 Quantum probability as a normative decision theory

What might a symplectic algorithm for solving problem **P** look like? I don't have a complete picture, but I'd like to point out what I believe will be a key element of the answer.

It is clear that we need a probabilistic framework to handle the incomplete information in problem **P**. However, the framework of *classical* probability doesn't seem to be well suited to exploit symplectic structure. To clarify what I mean, notice that the space of all valid classical probability assignments, for a given problem, is some n-dimensional simplex, Δ^n.[1] But for n even there's no useful way, as far as I know, to endow Δ^n with a symplectic structure; while for n odd, Δ^n doesn't even admit of a symplectic structure at all.[2] So there indeed seems to be a mismatch between the tool (classical probability) and the task (exploiting symplecticity). Now, we do know of a *non-classical* probabilistic framework that has a natural symplectic structure: it is a type of non-commutative probability [268]; specifically, that used in quantum mechanics [269]. In this framework, the space of all valid probability assignments, for a given problem, consists of the convex linear combinations of certain special elements called "pure states". Pure states form a space of their own: a $2n$-dimensional complex projective space, $\mathbb{CP}^n$. And indeed, $\mathbb{CP}^n$ is a symplectic manifold for any n [259].

These considerations suggest that a symplectic algorithm for solving problem **P** must rely on quantum probability, not classical probability, to quantify uncertainties. I take this suggestion one step further by proposing the following conjecture.

[1] Namely, for a random variable with $n+1$ possible values, valid probability assignments are in one-to-one correspondence with $(n+1)$-tuples of positive real numbers that sum to one: $\Delta^n = \{(p_1, \ldots, p_{n+1}) \,|\, p_1 + \cdots + p_{n+1} = 1\}$.

[2] Symplectic geometry requires an even-dimensional manifold.

Q: Perhaps there is a precise normative sense in which quantum probability is *the* rational way to quantify uncertainty in problem **P**.

A few clarifications are in order. (i) In decision theory [270, 271], probability theory is understood as a tool for making rational decisions under uncertainty.[3] Regarded in this way, the whole framework of (classical) probability theory comes about, not as a calculus of empirical frequencies, nor as the logical consequences of a set of axioms, but as practical normative rules; rules of conduct which we are free to violate, but we do so at our own peril by exposing ourselves to the possibility of sure losses. This is the sense in which I use the words "normative", "probability" and "rational" in conjecture **Q**. (ii) I haven't given a precise definition of "quantum probability". Just what part of the mathematics of quantum physics am I proposing will play a role here? Am I perhaps suggesting that the special issues of the particular relativistic quantum field theories appearing in the standard model of particle physics will be relevant? Or that Planck's constant $\hbar \approx 1 \times 10^{-34}$J·s will appear? Of course not. Something like the following seems appropriate: that propositions can be made to correspond to subspaces (equivalently, to orthogonal projection operators) of a Hilbert space, $\mathcal{H}$, over the field $\mathbb{C}$;[4] that real-valued functions over phase space can be made to correspond to hermitian operators over $\mathcal{H}$, in such a way that the Lie algebra of functions be homomorphic to the Lie algebra of operators, as in canonical quantization;[5] that probability assignments can be made to correspond to von Neumann density operators over $\mathcal{H}$ in the usual way; and that the optimal trajectories—the solutions to problem **P**—can be made to correspond to unitary evolution over $\mathcal{H}$. (iii) I'd like to emphasize that conjecture **Q** is a mathematical conjecture, not, say, a philosophical perspective or even a scientific hypothesis; it's either true or false; and what it would take to settle the conjecture in the positive is a Dutch Book theorem, analogous to those of classical decision theory [271], of the following form. Faced with "such-and-such a family of decision problems of the form **P**", any agent whose choice preferences do not meet "such-and-such conditions to do with quantum probability" is subject to "such-and-such a money pump" that guarantees a loss in every possible outcome.

Aside from the argument presented above involving the desire to exploit symplectic structure for algorithmic gains, further support for conjecture **Q** can be drawn from other places, as follows. (i) Consider the fact that quantum probability is required in physics for the microscopic description of systems, which macroscopically are very well described by Hamiltonian mechanics. This empirical observation has, of course, a theoretical counterpart, in Bohr's correspondence principle [274], and in the various known prescriptions for "taking the classical limit" of a quantum theory [273, 275, 276] and, in the other direction, prescriptions for "quantizing" a classical theory [273]. Said prescriptions always relate a quantum theory to a classical Hamiltonian theory. This correspondence speaks to a deep compatibility between variational problems such as (1) (and hence such as (2)) and the quantum probabilistic framework.

(ii) As we have seen in some detail in Part I, even in a hypothetical world governed by classical Hamiltonian mechanics, uncertainty regarding the Hamiltonian of a system seems to lead to uncertainty about the state of the system, which is of the same form as Heisenberg's uncertainty relation. In turn, Spekkens [41, 42] and Bartlett *et al.* [43] have shown that in classical ontologies where the epistemology is restricted by a Heisenberg-like uncertainty relation, many of the phenomena which were thought to be characteristic

[3] In decision theory, probabilities are, by definition, choice preferences. E.g. Let \$p be the highest price that You would be willing to pay now for a lottery ticket worth \$1 if it rains tomorrow, \$0 otherwise. We call p the probability, for You, that it will rain tomorrow.

[4] It's important that the field be $\mathbb{C}$—not $\mathbb{R}$ or $\mathbb{H}$—so that the Hilbert space be a symplectic manifold.

[5] I'm glossing over well-known issues to do with obstructions to quantization, such as the Groenewold-van Hove no-go theorem [272, 273].

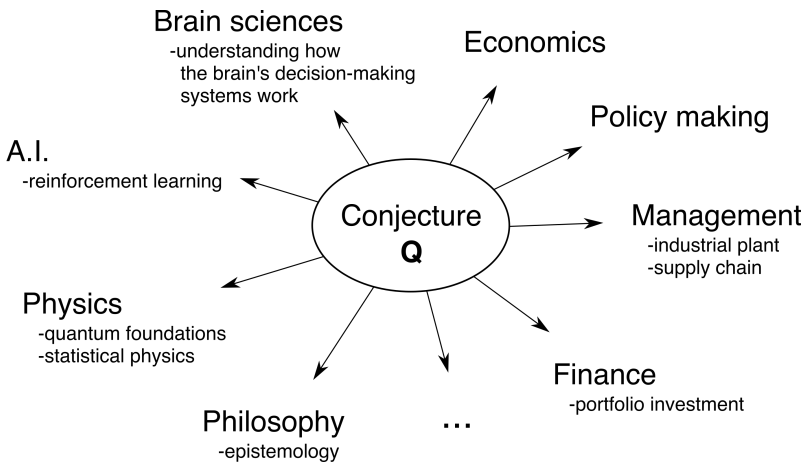


Figure 1: **Sphere of influence of a decision-theoretic result.** If correct, conjecture **Q** may have far-reaching consequences across the sciences and beyond.

of quantum mechanics can occur.[6] Taken together, these results support the idea that quantum probability is the natural way to quantify uncertainty in variational problems such as (1) and (2). "Natural" in the sense that it will straightforwardly yield valid solutions to problem **P** which would otherwise seem counterintuitive, or even paradoxical, if one carelessly forced classical probability on the problem.

(iii) An independent line of evidence comes from behavioral studies in cognitive science. Beginning in the 1970's with the work of Tversky and Kahneman [277–281], psychologists began to uncover decision tasks for which humans deviate systematically from classical decision theory. These are sometimes called "paradoxical behaviors". Examples are the Allais paradox [282]; the Ellsberg paradox [283]; question-order effects [284, 285]; the conjunction fallacy [286]; the disjunction fallacy [287, 288][289, p. 126]; and violations of Savage's sure-thing principle, e.g. violations of the law of total probability, such as the disjunction effect [290]. One approach to explain such paradoxical behaviors goes by the name of *quantum-like models of cognition* [289]. The idea is to substitute quantum probability, in place of classical probability, in otherwise classical decision models. The approach has been gaining popularity over the last decade. It is noteworthy for its parsimonious account of a substantial number paradoxical behaviors [289, 291–297]; and for successfully predicting one novel paradoxical effect with quantitative accuracy [298]. (Note that this was a zero-parameter, quantitative, a priori prediction, which is uncommon in social sciences.) The success of these descriptive models lends support to our conjecture, **Q**, by suggesting that the usage of quantum probability by certain brain system(s) has been selected for by evolution, which could only happen if quantum probability conferred a selective advantage over alternative approaches in some decision problem(s).

Finally, we should note that the idea is not new that quantum theory may enjoy some sort of normative status. It can be found at least as far back as two decades, in the literature on quantum foundations, specifically in ψ-epistemic interpretations [39, 299, 300] and in the many-worlds interpretation [301, 302]. (An example is the epigraph at the beginning of this Part III.) I do wonder if anybody has had in mind as mundane a decision problem as **P**.

[6]The list of phenomena includes noncommutativity, coherent superposition, collapse, complementarity, no-cloning, no-broadcasting, interference, teleportation, remote steering, key distribution, dense coding, entanglement, monogamy of entanglement, ambiguity of mixtures, locally immeasurable product bases, unextendible product bases, pre and post-selection effects, quantum eraser and many others.

III.4 A bigger picture

We've motivated problem **P** and conjecture **Q** from the point of view of artificial intelligence and reinforcement learning. But **P** is a quite general statement of the problem of intertemporal choice under uncertainty; so that conjecture **Q**, if it were to prove true, would be a general result in decision theory. As indicated in Figure 1, such a breakthrough would be consequential for many fields of human endeavor aside from A.I. Outside of academia: policy making, management and finance are just some of the endeavors informed by decision theory. In academia: cognitive science routinely draws from the normative results of decision theory, to inspire new computational models of decision-making in living organisms. Neuroscience cares about how such computations may be implemented by the central nervous system. And of course, intertemporal choice in humans is a central theme of economics. In philosophy, epistemology cares about the proper ways to quantify knowledge, and what their limits may be—both of which have been directly informed by Bayesian probability theory in the past, and would be so again if conjecture **Q** proved true. I believe that physics, too, would stand to gain. A positive answer to conjecture **Q** would be informative to debates around quantum foundations. Concretely, I can see it bolstering support for certain so-called "retrocausal", or "all-at-once", ψ-epistemic interpretations of quantum mechanics [303].

Bibliography

1. Ozawa, M. Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement. *Phys. Rev. A* **67,** 042105 (2003).
2. Heisenberg, W. *The physical principles of the quantum theory* (Dover, 1949).
3. Lamb, W. E. & Fearn, H. in *Amazing Light* 373–389 (Springer, 1996).
4. Morgan, P. An algebraic approach to Koopman classical mechanics. *Ann. Phys.* **414,** 168090 (2020).
5. Katagiri, S. Measurement theory in classical mechanics. *Prog. Theor. Exp. Phys.* **2020,** 063A02 (2020).
6. Collier, J. Two faces of Maxwell's demon reveal the nature of irreversibility. *Stud. Hist. Philos. Sci.* **21,** 22J (1990).
7. Szilard, L. Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Z. Phys.* **53,** 840–856 (1929).
8. Von Neumann, J. *Mathematical foundations of quantum mechanics* Original work published 1932 (Princeton University Press, 1955).
9. Brillouin, L. Maxwell's demon cannot operate: Information and entropy. I. *J. Appl. Phys.* **22,** 334–337 (1951).
10. Brillouin, L. The negentropy principle of information. *J. Appl. Phys.* **24,** 1152–1163 (1953).
11. Gabor, D. §5 A Further Paradox: A Perpetuum Mobile of the Second Kind. *Reprinted in Leff and Rex (1990),* 148–159 (1951).
12. Bennett, C. H. The thermodynamics of computation—a review. *Int. J. Theor. Phys.* **21,** 905–940 (1982).
13. Bennett, C. H. Demons, engines and the second law. *Sci. Am.* **257,** 108–117 (1987).
14. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5,** 183–191 (1961).
15. Sagawa, T. Thermodynamics of information processing in small systems. *Prog. Theor. Phys.* **127,** 1–56 (2012).
16. Jacobs, K. & Steck, D. A. A straightforward introduction to continuous quantum measurement. *Contemp. Phys.* **47,** 279–303 (2006).
17. Arnol'd, V. I. *Mathematical methods of classical mechanics* 2nd ed. (Springer, 1989).
18. Libermann, P. & Marle, C.-M. *Symplectic geometry and analytical mechanics* (Springer Science & Business Media, 2012).
19. Whittaker, E. T. *A treatise on the analytical dynamics of particles and rigid bodies* 4th ed. section 192 (Cambridge University Press, London, 1959).
20. Williamson, J. On the algebraic problem concerning the normal forms of linear dynamical systems. *Am. J. Math.* **58,** 141–163 (1936).

21. Kardar, M. *Statistical physics of particles* (Cambridge University Press, 2007).

22. Bismut, J.-M. *Mécanique aléatoire* (Springer-Verlag, Berlin, 1981).

23. Parrondo, J. M. R., Mañas, M. & de la Rubia, F. J. Geometrical treatment of systems driven by coloured noise. *J. Phys. A-Math Gen* **23,** 2363 (1990).

24. Arthurs, E. & Kelly, J. L. BSTJ Briefs: On the simultaneous measurement of a pair of conjugate observables. *Bell Syst. Tech. J.* **44,** 725–729 (1965).

25. Arthurs, E. & Goodman, M. S. Quantum correlations: A generalized Heisenberg uncertainty relation. *Phys. Rev. Lett.* **60,** 2447 (1988).

26. Ozawa, M. in *Quantum Aspects of Optical Communications* (eds Bendjaballah, C., Hirota, O. & Reynaud, S.) 1–17 (Springer, 1991).

27. Ishikawa, S. Uncertainty relations in simultaneous measurements for arbitrary observables. *Rep. Math. Phys.* **29,** 257–273 (1991).

28. Erhart, J. *et al.* Experimental demonstration of a universally valid error–disturbance uncertainty relation in spin measurements. *Nat. Phys.* **8,** 185–189 (2012).

29. Fujikawa, K. Universally valid Heisenberg uncertainty relation. *Phys. Rev. A* **85,** 062117 (2012).

30. Baek, S.-Y., Kaneda, F., Ozawa, M. & Edamatsu, K. Experimental violation and reformulation of the Heisenberg's error-disturbance uncertainty relation. *Sci. Rep.* **3,** 2221 (2013).

31. Busch, P., Lahti, P. & Werner, R. F. Proof of Heisenberg's error-disturbance relation. *Phys. Rev. Lett.* **111,** 160405 (2013).

32. Griffiths, D. J. *Introduction to quantum mechanics* 2nd ed. (Pearson Prentice Hall, 2005).

33. Wigner, E. P. in *The Scientist Speculates* (ed Good, I. J.) (Heinemann, London, 1961).

34. Bell, J. Against 'measurement'. *Phys. World* **3,** 33 (1990).

35. Nielsen, M. A. & Chuang, I. *Quantum computation and quantum information* (American Association of Physics Teachers, 2002).

36. Wiseman, H. Quantum trajectories and quantum measurement theory. *Quantum Semicl. Opt.* **8,** 205 (1996).

37. Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47,** 777 (1935).

38. Harrigan, N. & Spekkens, R. W. Einstein, incompleteness, and the epistemic view of quantum states. *Found. Phys.* **40,** 125–157 (2010).

39. Fuchs, C. A. *Quantum mechanics as quantum information (and only a little more)* 2002. arXiv: `quant-ph/0205039`.

40. Caves, C. M., Fuchs, C. A. & Schack, R. Unknown quantum states: the quantum de Finetti representation. *J. Math. Phys.* **43,** 4537–4559 (2002).

41. Spekkens, R. W. Evidence for the epistemic view of quantum states: A toy theory. *Phys. Rev. A* **75,** 032110 (2007).

42. Spekkens, R. W. in *Quantum Theory: Informational Foundations and Foils* 83–135 (Springer, 2016).

43. Bartlett, S. D., Rudolph, T. & Spekkens, R. W. Reconstruction of Gaussian quantum mechanics from Liouville mechanics with an epistemic restriction. *Phys. Rev. A* **86,** 012103 (2012).

44. Frank, M. P. *Foundations of generalized reversible computing* in *International Conference on Reversible Computation* (2017), 19–34.

45. Arnol'd, V. I. & Givental', A. B. in *Dynamical Systems IV: Symplectic geometry and its applications* 1–136 (Berlin: Springer-Verlag, 1990).

46. Kirk, D. E. *Optimal control theory: an introduction* (Prentice-Hall, Inc., 1970).

47. Cohen, N. J. & Eichenbaum, H. *Memory, amnesia, and the hippocampal system* (MIT press, 1995).

48. O'Keefe, J. & Nadel, L. *The hippocampus as a cognitive map* (Oxford: Clarendon Press, 1978).

49. Redish, A. D. *et al. Beyond the cognitive map: from place cells to episodic memory* (MIT press, 1999).

50. Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M. & Tanila, H. The hippocampus, memory, and place cells: is it spatial memory or a memory space? *Neuron* **23,** 209–226 (1999).

51. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annual review of neuroscience* **24,** 167–202 (2001).

52. Colgin, L. L. Oscillations and hippocampal–prefrontal synchrony. *Current opinion in neurobiology* **21,** 467–474 (2011).

53. Preston, A. R. & Eichenbaum, H. Interplay of hippocampus and prefrontal cortex in memory. *Current Biology* **23,** R764–R773 (2013).

54. Wang, J. X., Cohen, N. J. & Voss, J. L. Covert rapid action-memory simulation (CRAMS): A hypothesis of hippocampal–prefrontal interactions for adaptive behavior. *Neurobiology of learning and memory* **117,** 22–33 (2015).

55. Redish, A. D. Vicarious trial and error. *Nature Reviews Neuroscience* **17,** 147 (2016).

56. Penagos, H., Varela, C. & Wilson, M. A. Oscillations, neural computations and learning during wake and sleep. *Current opinion in neurobiology* **44,** 193–201 (2017).

57. Eichenbaum, H. Prefrontal–hippocampal interactions in episodic memory. *Nature Reviews Neuroscience* **18,** 547 (2017).

58. Van der Meer, M., Kurth-Nelson, Z. & Redish, A. D. Information processing in decision-making systems. *The Neuroscientist* **18,** 342–359 (2012).

59. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80,** 312–325 (2013).

60. Marr, D. & Poggio, T. From understanding computation to understanding neural circuitry. *MIT, A.I. Laboratory Memos* (1976).

61. Dayan, P. Improving generalization for temporal difference learning: The successor representation. *Neural Computation* **5,** 613–624 (1993).

62. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nature neuroscience* **20,** 1643 (2017).

63. Mattar, M. G. & Daw, N. D. Prioritized memory access explains planning and hippocampal replay. *Nature neuroscience* **21,** 1609 (2018).

64. Buzsáki, G. Theta oscillations in the hippocampus. *Neuron* **33,** 325–340 (2002).

65. Foster, D. J. & Wilson, M. A. Hippocampal theta sequences. *Hippocampus* **17,** 1093–1099 (2007).

66. Buzsáki, G. Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus* **25,** 1073–1188 (2015).

67. Foster, D. J. Replay comes of age. *Annual review of neuroscience* **40,** 581–602 (2017).

68. Feng, T., Silva, D. & Foster, D. J. Dissociation between the experience-dependent development of hippocampal theta sequences and single-trial phase precession. *Journal of Neuroscience* **35,** 4890–4902 (2015).

69. Davidson, T. J., Kloosterman, F. & Wilson, M. A. Hippocampal replay of extended experience. *Neuron* **63,** 497–507 (2009).

70. Karlsson, M. P. & Frank, L. M. Awake replay of remote experiences in the hippocampus. *Nature neuroscience* **12,** 913–918 (2009).

71. Wikenheiser, A. M. & Redish, A. D. The balance of forward and backward hippocampal sequences shifts across behavioral states. *Hippocampus* **23,** 22–29 (2013).

72. Foster, D. J. & Wilson, M. A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440,** 680 (2006).

73. Buzsáki, G. Two-stage model of memory trace formation: a role for "noisy" brain states. *Neuroscience* **31,** 551–570 (1989).

74. Kay, K. *et al.* Constant sub-second cycling between representations of possible futures in the hippocampus. *bioRxiv,* 528976 (2019).

75. Carr, M. F., Jadhav, S. P. & Frank, L. M. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature neuroscience* **14,** 147 (2011).

76. Diba, K. & Buzsáki, G. Forward and reverse hippocampal place-cell sequences during ripples. *Nature neuroscience* **10,** 1241 (2007).

77. Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6,** 149–172 (1996).

78. Dragoi, G. & Buzsáki, G. Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron* **50,** 145–157 (2006).

79. Jadhav, S. P., Kemere, C., German, P. W. & Frank, L. M. Awake hippocampal sharp-wave ripples support spatial memory. *Science* **336,** 1454–1458 (2012).

80. Liu, K., Sibille, J. & Dragoi, G. Generative predictive codes by multiplexed hippocampal neuronal tuplets. *Neuron* **99,** 1329–1341 (2018).

81. Nicola, W. & Clopath, C. A diversity of interneurons and Hebbian plasticity facilitate rapid compressible learning in the hippocampus. *Nature Neuroscience* **22,** 1168 (2019).

82. Qin, Y.-L., McNaughton, B. L., Skaggs, W. E. & Barnes, C. A. Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **352,** 1525–1533 (1997).

83. Wagner, U., Gais, S., Haider, H., Verleger, R. & Born, J. Sleep inspires insight. *Nature* **427,** 352 (2004).

84. Ji, D. & Wilson, M. A. Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature neuroscience* **10,** 100 (2007).

85. Euston, D. R., Tatsuno, M. & McNaughton, B. L. Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *science* **318,** 1147–1150 (2007).

86. Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I. & Battaglia, F. P. Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature neuroscience* **12,** 919 (2009).

87. Lansink, C. S., Goltstein, P. M., Lankelma, J. V., McNaughton, B. L. & Pennartz, C. M. Hippocampus leads ventral striatum in replay of place-reward information. *PLoS biology* **7,** e1000173 (2009).

88. Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G. & Zugaro, M. B. Selective suppression of hippocampal ripples impairs spatial memory. *Nature neuroscience* **12,** 1222 (2009).

89. Ego-Stengel, V. & Wilson, M. A. Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus* **20,** 1–10 (2010).

90. Born, J. & Wilhelm, I. System consolidation of memory during sleep. *Psychological research* **76,** 192–203 (2012).

91. Gupta, A. S., van der Meer, M. A., Touretzky, D. S. & Redish, A. D. Hippocampal replay is not a simple function of experience. *Neuron* **65,** 695–705 (2010).

92. Pezzulo, G., van der Meer, M. A., Lansink, C. S. & Pennartz, C. M. Internally generated sequences in learning and executing goal-directed behavior. *Trends in cognitive sciences* **18,** 647–657 (2014).

93. Jadhav, S. P., Rothschild, G., Roumis, D. K. & Frank, L. M. Coordinated excitation and inhibition of prefrontal ensembles during awake hippocampal sharp-wave ripple events. *Neuron* **90,** 113–127 (2016).

94. Fernández-Ruiz, A. *et al.* Long-duration hippocampal sharp wave ripples improve memory. *Science* **364,** 1082–1086 (2019).

95. Foster, D., Morris, R. & Dayan, P. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* **10,** 1–16 (2000).

96. Johnson, A. & Redish, A. D. Hippocampal replay contributes to within session learning in a temporal difference reinforcement learning model. *Neural Networks* **18,** 1163–1171 (2005).

97. Singer, A. C. & Frank, L. M. Rewarded outcomes enhance reactivation of experience in the hippocampus. *Neuron* **64,** 910–921 (2009).

98. Ambrose, R. E., Pfeiffer, B. E. & Foster, D. J. Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron* **91,** 1124–1136 (2016).

99. Momennejad, I. *et al.* The successor representation in human reinforcement learning. *Nature Human Behaviour* **1,** 680 (2017).

100. Johnson, A. & Redish, A. D. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience* **27,** 12176–12189 (2007).

101. Lisman, J. & Redish, A. D. Prediction, sequences and the hippocampus. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364,** 1193–1201 (2009).

102. Chersi, F. & Pezzulo, G. Using hippocampal-striatal loops for spatial navigation and goal-directed decision-making. *Cognitive processing* **13,** 125–129 (2012).

103. Chersi, F., Donnarumma, F. & Pezzulo, G. Mental imagery in the navigation domain: a computational model of sensory-motor simulation mechanisms. *Adaptive Behavior* **21,** 251–262 (2013).

104. Pezzulo, G., Rigoli, F. & Chersi, F. The mixed instrumental controller: using value of information to combine habitual choice and mental simulation. *Frontiers in psychology* **4,** 92 (2013).

105. Pfeiffer, B. E. & Foster, D. J. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497,** 74 (2013).

106. Wikenheiser, A. M. & Redish, A. D. Hippocampal theta sequences reflect current goals. *Nature neuroscience* **18,** 289 (2015).

107. Floresco, S. B., Seamans, J. K. & Phillips, A. G. Selective roles for hippocampal, prefrontal cortical, and ventral striatal circuits in radial-arm maze tasks with or without a delay. *Journal of Neuroscience* **17,** 1880–1890 (1997).

108. Hannesson, D. K., Howland, J. G. & Phillips, A. G. Interaction between perirhinal and medial prefrontal cortex is required for temporal order but not recognition memory for objects in rats. *Journal of Neuroscience* **24,** 4596–4604 (2004).

109. Wang, G.-W. & Cai, J.-X. Disconnection of the hippocampal–prefrontal cortical circuits impairs spatial working memory performance in rats. *Behavioural brain research* **175,** 329–336 (2006).

110. Barker, G. R., Bird, F., Alexander, V. & Warburton, E. C. Recognition memory for objects, place, and temporal order: a disconnection analysis of the role of the medial prefrontal cortex and perirhinal cortex. *Journal of Neuroscience* **27,** 2948–2957 (2007).

111. Barker, G. R. *et al.* Separate elements of episodic memory subserved by distinct hippocampal–prefrontal connections. *Nature neuroscience* **20,** 242 (2017).

112. Jones, M. W. & Wilson, M. A. Phase precession of medial prefrontal cortical activity relative to the hippocampal theta rhythm. *Hippocampus* **15,** 867–873 (2005).

113. Siapas, A. G., Lubenov, E. V. & Wilson, M. A. Prefrontal phase locking to hippocampal theta oscillations. *Neuron* **46,** 141–151 (2005).

114. Hyman, J. M., Zilli, E. A., Paley, A. M. & Hasselmo, M. E. Medial prefrontal cortex cells show dynamic modulation with the hippocampal theta rhythm dependent on behavior. *Hippocampus* **15,** 739–749 (2005).

115. Sirota, A. *et al.* Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. *Neuron* **60,** 683–697 (2008).

116. Paz, R., Bauer, E. P. & Paré, D. Theta synchronizes the activity of medial prefrontal neurons during learning. *Learning & Memory* **15,** 524–531 (2008).

117. Benchenane, K. *et al.* Coherent theta oscillations and reorganization of spike timing in the hippocampal-prefrontal network upon learning. *Neuron* **66,** 921–936 (2010).

118. Kim, J., Delcasso, S. & Lee, I. Neural correlates of object-in-place learning in hippocampus and prefrontal cortex. *Journal of Neuroscience* **31,** 16991–17006 (2011).

119. Backus, A. R., Schoffelen, J.-M., Szebényi, S., Hanslmayr, S. & Doeller, C. F. Hippocampal-prefrontal theta oscillations support memory integration. *Current Biology* **26,** 450–457 (2016).

120. Hallock, H. L., Wang, A. & Griffin, A. L. Ventral midline thalamus is critical for hippocampal–prefrontal synchrony and spatial working memory. *Journal of Neuroscience* **36,** 8372–8389 (2016).

121. Jones, M. W. & Wilson, M. A. Theta rhythms coordinate hippocampal–prefrontal interactions in a spatial memory task. *PLoS biology* **3,** e402 (2005).

122. Sigurdsson, T., Stark, K. L., Karayiorgou, M., Gogos, J. A. & Gordon, J. A. Impaired hippocampal–prefrontal synchrony in a genetic mouse model of schizophrenia. *Nature* **464,** 763 (2010).

123. Hyman, J. M., Zilli, E. A., Paley, A. M. & Hasselmo, M. E. Working memory performance correlates with prefrontal-hippocampal theta interactions but not with prefrontal neuron firing rates. *Frontiers in integrative neuroscience* **4,** 2 (2010).

124. O'Neill, P.-K., Gordon, J. A. & Sigurdsson, T. Theta oscillations in the medial prefrontal cortex are modulated by spatial working memory and synchronize with the hippocampus through its ventral subregion. *Journal of Neuroscience* **33,** 14211–14224 (2013).

125. Spellman, T. *et al.* Hippocampal–prefrontal input supports spatial encoding in working memory. *Nature* **522,** 309 (2015).

126. Place, R., Farovik, A., Brockmann, M. & Eichenbaum, H. Bidirectional prefrontal-hippocampal interactions support context-guided memory. *Nature neuroscience* **19,** 992 (2016).

127. Adhikari, A., Topiwala, M. A. & Gordon, J. A. Synchronized activity between the ventral hippocampus and the medial prefrontal cortex during anxiety. *Neuron* **65,** 257–269 (2010).

128. Steriade, M. Grouping of brain rhythms in corticothalamic systems. *Neuroscience* **137,** 1087–1106 (2006).

129. Neske, G. T. The slow oscillation in cortical and thalamic networks: mechanisms and functions. *Frontiers in neural circuits* **9,** 88 (2016).

130. Staresina, B. P. *et al.* Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature neuroscience* **18,** 1679 (2015).

131. Varela, C. & Wilson, M. A. mPFC spindle cycles organize sparse thalamic activation and recently active CA1 cells during non-REM sleep. *Elife* **9,** e48881 (2020).

132. Destexhe, A., Hughes, S. W., Rudolph, M. & Crunelli, V. Are corticothalamic 'up' states fragments of wakefulness? *Trends in neurosciences* **30,** 334–342 (2007).

133. Russell, S. J. & Norvig, P. *Artificial intelligence: a modern approach* (Malaysia; Pearson Education Limited, 2016).

134. Holland, P. C. & Bouton, M. E. Hippocampus and context in classical conditioning. *Current opinion in neurobiology* **9,** 195–202 (1999).

135. Corcoran, K. A. & Maren, S. Hippocampal inactivation disrupts contextual retrieval of fear memory after extinction. *Journal of Neuroscience* **21,** 1720–1726 (2001).

136. Eichenbaum, H. Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron* **44,** 109–120 (2004).

137. Eacott, M. J. & Norman, G. Integrated memory for object, place, and context in rats: a possible model of episodic-like memory? *Journal of Neuroscience* **24,** 1948–1953 (2004).

138. Langston, R. F. & Wood, E. R. Associative recognition and the hippocampus: Differential effects of hippocampal lesions on object-place, object-context and object-place-context memory. *Hippocampus* **20,** 1139–1153 (2010).

139. Butterly, D. A., Petroccione, M. A. & Smith, D. M. Hippocampal context processing is critical for interference free recall of odor memories in rats. *Hippocampus* **22,** 906–913 (2012).

140. Eichenbaum, H. Memory: organization and control. *Annual review of psychology* **68,** 19–45 (2017).

141. Moita, M. A., Rosis, S., Zhou, Y., LeDoux, J. E. & Blair, H. T. Hippocampal place cells acquire location-specific responses to the conditioned stimulus during auditory fear conditioning. *Neuron* **37,** 485–497 (2003).

142. Manns, J. R. & Eichenbaum, H. A cognitive map for object memory in the hippocampus. *Learning & Memory* **16,** 616–624 (2009).

143. Komorowski, R. W., Manns, J. R. & Eichenbaum, H. Robust conjunctive item–place coding by hippocampal neurons parallels learning what happens where. *Journal of Neuroscience* **29,** 9918–9929 (2009).

144. Itskov, P. M., Vinnik, E. & Diamond, M. E. Hippocampal representation of touch-guided behavior in rats: persistent and independent traces of stimulus and reward location. *PloS one* **6,** e16462 (2011).

145. Itskov, P. M., Vinnik, E., Honey, C., Schnupp, J. & Diamond, M. E. Sound sensitivity of neurons in rat hippocampus during performance of a sound-guided task. *Journal of neurophysiology* **107,** 1822–1834 (2012).

146. Vinnik, E., Antopolskiy, S., Itskov, P. M. & Diamond, M. E. Auditory stimuli elicit hippocampal neuronal responses during sleep. *Frontiers in systems neuroscience* **6,** 49 (2012).

147. MacDonald, C. J., Carrow, S., Place, R. & Eichenbaum, H. Distinct hippocampal time cell sequences represent odor memories in immobilized rats. *Journal of Neuroscience* **33,** 14607–14616 (2013).

148. Bulkin, D. A., Law, L. M. & Smith, D. M. Placing memories in context: Hippocampal representations promote retrieval of appropriate memories. *Hippocampus* **26,** 958–971 (2016).

149. Igarashi, K. M., Ito, H. T., Moser, E. I. & Moser, M.-B. Functional diversity along the transverse axis of hippocampal area CA1. *FEBS letters* **588,** 2470–2476 (2014).

150. Kjelstrup, K. B. *et al.* Finite scale of spatial representation in the hippocampus. *Science* **321,** 140–143 (2008).

151. Royer, S., Sirota, A., Patel, J. & Buzsáki, G. Distinct representations and theta dynamics in dorsal and ventral hippocampus. *Journal of Neuroscience* **30,** 1777–1787 (2010).

152. Fanselow, M. S. & Dong, H.-W. Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron* **65,** 7–19 (2010).

153. Komorowski, R. W. *et al.* Ventral hippocampal neurons are shaped by experience to represent behaviorally relevant contexts. *Journal of Neuroscience* **33,** 8079–8087 (2013).

154. Poppenk, J., Evensmoen, H. R., Moscovitch, M. & Nadel, L. Long-axis specialization of the human hippocampus. *Trends in cognitive sciences* **17,** 230–240 (2013).

155. Strange, B. A., Witter, M. P., Lein, E. S. & Moser, E. I. Functional organization of the hippocampal longitudinal axis. *Nature Reviews Neuroscience* **15,** 655–669 (2014).

156. Czerniawski, J., Yoon, T. & Otto, T. Dissociating space and trace in dorsal and ventral hippocampus. *Hippocampus* **19,** 20–32 (2009).

157. Moscovitch, M. Memory and working-with-memory: A component process model based on modules and central systems. *Journal of cognitive neuroscience* **4,** 257–267 (1992).

158. Dobbins, I. G., Foley, H., Schacter, D. L. & Wagner, A. D. Executive control during episodic retrieval: multiple prefrontal processes subserve source memory. *Neuron* **35,** 989–996 (2002).

159. Postle, B. R. Working memory as an emergent property of the mind and brain. *Neuroscience* **139,** 23–38 (2006).

160. Ranganath, C. & Blumenfeld, R. S. in *Learning and Memory: A Comprehensive Reference* 1st ed., 261–279 (Oxford Univ. Press, 2008).

161. Kuhl, B. A. & Wagner, A. D. in *Encyclopedia of Neuroscience* 437–444 (Oxford: Academic Press, 2009).

162. Szczepanski, S. M. & Knight, R. T. Insights into human behavior from lesions to the prefrontal cortex. *Neuron* **83,** 1002–1018 (2014).

163. Rich, E. L. & Shapiro, M. Rat prefrontal cortical neurons selectively code strategy switches. *Journal of Neuroscience* **29,** 7208–7219 (2009).

164. Durstewitz, D., Vittoz, N. M., Floresco, S. B. & Seamans, J. K. Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* **66,** 438–448 (2010).

165. Karlsson, M. P., Tervo, D. G. & Karpova, A. Y. Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* **338,** 135–139 (2012).

166. Ma, L., Hyman, J. M., Durstewitz, D., Phillips, A. G. & Seamans, J. K. A quantitative analysis of context-dependent remapping of medial frontal cortex neurons and ensembles. *Journal of Neuroscience* **36,** 8258–8272 (2016).

167. Guise, K. G. & Shapiro, M. L. Medial prefrontal cortex reduces memory interference by modifying hippocampal encoding. *Neuron* **94,** 183–192 (2017).

168. Morrissey, M. D., Insel, N. & Takehara-Nishiuchi, K. Generalizable knowledge outweighs incidental details in prefrontal ensemble code over time. *Elife* **6,** e22177 (2017).

169. Miller, E. K., Freedman, D. J. & Wallis, J. D. The prefrontal cortex: categories, concepts and cognition. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **357,** 1123–1136 (2002).

170. Buschman, T. J., Denovellis, E. L., Diogo, C., Bullock, D. & Miller, E. K. Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron* **76,** 838–846 (2012).

171. Stokes, M. G. *et al.* Dynamic coding for cognitive control in prefrontal cortex. *Neuron* **78,** 364–375 (2013).

172. Blackman, R. K. *et al.* Monkey prefrontal neurons reflect logical operations for cognitive control in a variant of the AX continuous performance task (AX-CPT). *Journal of Neuroscience* **36,** 4067–4079 (2016).

173. Botvinick, M. M., Niv, Y. & Barto, A. C. Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition* **113,** 262–280 (2009).

174. Holroyd, C. B. & Yeung, N. Motivation of extended behaviors by anterior cingulate cortex. *Trends in cognitive sciences* **16,** 122–128 (2012).

175. Ragozzino, M. E., Detrick, S. & Kesner, R. P. Involvement of the prelimbic–infralimbic areas of the rodent prefrontal cortex in behavioral flexibility for place and response learning. *Journal of Neuroscience* **19,** 4585–4594 (1999).

176. Wallis, J. D., Anderson, K. C. & Miller, E. K. Single neurons in prefrontal cortex encode abstract rules. *Nature* **411,** 953 (2001).

177. Brown, V. J. & Bowman, E. M. Rodent models of prefrontal cortical function. *Trends in neurosciences* **25,** 340–343 (2002).

178. Ragozzino, M. E., Kim, J., Hassert, D., Minniti, N. & Kiang, C. The contribution of the rat prelimbic-infralimbic areas to different forms of task switching. *Behavioral neuroscience* **117,** 1054 (2003).

179. Mulder, A. B., Nordquist, R. E., Örgüt, O. & Pennartz, C. M. Learning-related changes in response patterns of prefrontal neurons during instrumental conditioning. *Behavioural brain research* **146,** 77–88 (2003).

180. Hok, V., Save, E., Lenck-Santini, P. & Poucet, B. Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex. *Proceedings of the National Academy of Sciences* **102,** 4602–4607 (2005).

181. Marquis, J.-P., Killcross, S. & Haddon, J. E. Inactivation of the prelimbic, but not infralimbic, prefrontal cortex impairs the contextual control of response conflict in rats. *European Journal of Neuroscience* **25,** 559–566 (2007).

182. Rich, E. L. & Shapiro, M. L. Prelimbic/infralimbic inactivation impairs memory for multiple task switches, but not flexible selection of familiar tasks. *Journal of Neuroscience* **27,** 4747–4755 (2007).

183. Kehagia, A. A., Murray, G. K. & Robbins, T. W. Learning and cognitive flexibility: frontostriatal function and monoaminergic modulation. *Current opinion in neurobiology* **20,** 199–204 (2010).

184. Euston, D. R., Gruber, A. J. & McNaughton, B. L. The role of medial prefrontal cortex in memory and decision making. *Neuron* **76,** 1057–1070 (2012).

185. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503,** 78 (2013).

186. Vertes, R. P. Interactions among the medial prefrontal cortex, hippocampus and midline thalamus in emotional and cognitive processing in the rat. *Neuroscience* **142,** 1–20 (2006).

187. Hoover, W. B. & Vertes, R. P. Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Structure and Function* **212,** 149–179 (2007).

188. Jay, T. M., Glowinski, J. & Thierry, A.-M. Selectivity of the hippocampal projection to the prelimbic area of the prefrontal cortex in the rat. *Brain research* **505,** 337–340 (1989).

189. Jay, T. M. & Witter, M. P. Distribution of hippocampal CA1 and subicular efferents in the prefrontal cortex of the rat studied by means of anterograde transport of Phaseolus vulgaris-leucoagglutinin. *Journal of Comparative Neurology* **313,** 574–586 (1991).

190. Vertes, R. P., Hoover, W. B., Szigeti-Buck, K. & Leranth, C. Nucleus reuniens of the midline thalamus: link between the medial prefrontal cortex and the hippocampus. *Brain research bulletin* **71,** 601–609 (2007).

191. Cassel, J.-C. *et al.* The reuniens and rhomboid nuclei: neuroanatomy, electrophysiological characteristics and behavioral implications. *Progress in neurobiology* **111,** 34–52 (2013).

192. Wouterlood, F. G., Saldana, E. & Witter, M. P. Projection from the nucleus reuniens thalami to the hippocampal region: light and electron microscopic tracing study in the rat with the anterograde tracer Phaseolus vulgaris-leucoagglutinin. *Journal of Comparative Neurology* **296,** 179–203 (1990).

193. Dolleman-van Der Weel, M. & Witter, M. Projections from the nucleus reuniens thalami to the entorhinal cortex, hippocampal field CA1, and the subiculum in the rat arise from different populations of neurons. *Journal of comparative neurology* **364,** 637–650 (1996).

194. Vertes, R. P. Analysis of projections from the medial prefrontal cortex to the thalamus in the rat, with emphasis on nucleus reuniens. *Journal of Comparative Neurology* **442,** 163–187 (2002).

195. Varela, C., Kumar, S., Yang, J. & Wilson, M. Anatomical substrates for direct interactions between hippocampus, medial prefrontal cortex, and the thalamic nucleus reuniens. *Brain Structure and Function* **219,** 911–929 (2014).

196. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436,** 801 (2005).

197. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. Path integration and the neural basis of the'cognitive map'. *Nature Reviews Neuroscience* **7,** 663 (2006).

198. Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLoS computational biology* **5,** e1000291 (2009).

199. Insausti, R. & Amaral, D. G. in *The Human Nervous System* (eds Paxinos, G. & Mai, J. K.) 871–914 (Elsevier Academic Press, San Diego, CA, 2004).

200. Van Strien, N., Cappaert, N. & Witter, M. The anatomy of memory: an interactive overview of the parahippocampal–hippocampal network. *Nature Reviews Neuroscience* **10,** 272 (2009).

201. Cappaert, N. L., Van Strien, N. M. & Witter, M. P. in *The Rat Nervous System* (ed Paxinos, G.) 4th ed., 511–573 (Elsevier, 2015).

202. Ito, H. T., Zhang, S.-J., Witter, M. P., Moser, E. I. & Moser, M.-B. A prefrontal–thalamo–hippocampal circuit for goal-directed spatial navigation. *Nature* **522,** 50 (2015).

203. Brass, M. & Haggard, P. To do or not to do: the neural signature of self-control. *Journal of Neuroscience* **27,** 9141–9145 (2007).

204. Haggard, P. Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience* **9,** 934 (2008).

205. Verbruggen, F. & Logan, G. D. Response inhibition in the stop-signal paradigm. *Trends in cognitive sciences* **12,** 418–424 (2008).

206. Kühn, S., Haggard, P. & Brass, M. Intentional inhibition: How the "veto-area" exerts control. *Human brain mapping* **30,** 2834–2843 (2009).

207. Rae, C. L., Hughes, L. E., Anderson, M. C. & Rowe, J. B. The prefrontal cortex achieves inhibitory control by facilitating subcortical motor pathway connectivity. *Journal of neuroscience* **35,** 786–794 (2015).

208. Korb, S., Goldman, R., Davidson, R. J. & Niedenthal, P. M. Increased medial prefrontal cortex and decreased zygomaticus activation in response to disliked smiles suggest top-down inhibition of facial mimicry. *Frontiers in psychology* **10,** 1715 (2019).

209. Kolb, B. E. & Tees, R. C. *The cerebral cortex of the rat.* (The MIT Press, 1990).

210. Dalley, J. W., Cardinal, R. N. & Robbins, T. W. Prefrontal executive and cognitive functions in rodents: neural and neurochemical substrates. *Neuroscience & Biobehavioral Reviews* **28,** 771–784 (2004).

211. Castañeda, C. *The teachings of Don Juan: A Yaqui way of knowledge* (Univ of California Press, 1998).

212. Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction* 2nd ed. (MIT press, 2018).

213. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).

214. Ludvig, E. A., Bellemare, M. G. & Pearson, K. G. in *Computational neuroscience for advancing artificial intelligence: Models, methods and applications* 111–144 (IGI Global, 2011).

215. Shah, A. in *Reinforcement Learning* 507–537 (Springer, 2012).

216. Niv, Y., Joel, D. & Dayan, P. A normative perspective on motivation. *Trends in cognitive sciences* **10,** 375–381 (2006).

217. Dickinson, A. Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **308,** 67–78 (1985).

218. Devadoss, S. L. & O'Rourke, J. *Discrete and computational geometry* (Princeton University Press, 2011).

219. Schättler, H. & Ledzewicz, U. *Geometric optimal control: theory, methods and examples* (Springer, 2012).

220. Greif, C. *Numerical Methods for Hamilton-Jacobi-Bellman Equations* MA thesis (University of Wisconsin Milwaukee, 2017).

221. McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review* **102,** 419 (1995).

222. Miller, G. A., Eugene, G. & Pribram, K. H. *Plans and the Structure of Behaviour* (Henry Holt and Company, 1960).

223. Badre, D. & D'esposito, M. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience* **10,** 659–669 (2009).

224. Mehta, M. R., Quirk, M. C. & Wilson, M. A. Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* **25,** 707–715 (2000).

225. Low, R. J., Lewallen, S., Aronov, D., Nevers, R. & Tank, D. W. Probing variability in a cognitive map using manifold inference from neural dynamics. *BioRxiv,* 418939 (2018).

226. Deacon, R. M. & Rawlins, J. N. P. T-maze alternation in the rodent. *Nature protocols* **1,** 7–12 (2006).

227. McNaughton, B. L. *et al.* Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *The Journal of experimental biology* **199,** 173–185 (1996).

228. Redish, A. D. The hippocampal debate: are we asking the right questions? *Behavioural brain research* **127,** 81–98 (2001).

229. Frank, L. M., Brown, E. N. & Wilson, M. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* **27,** 169–178 (2000).

230. Griffin, A. L. & Hallock, H. L. Hippocampal signatures of episodic memory: evidence from single-unit recording studies. *Frontiers in behavioral neuroscience* **7,** 54 (2013).

231. Wood, E. R., Dudchenko, P. A., Robitsek, R. J. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27,** 623–633 (2000).

232. Ferbinteanu, J. & Shapiro, M. L. Prospective and retrospective memory coding in the hippocampus. *Neuron* **40,** 1227–1239 (2003).

233. Ainge, J. A., Tamosiunaite, M., Woergoetter, F. & Dudchenko, P. A. Hippocampal CA1 place cells encode intended destination on a maze with multiple choice points. *Journal of Neuroscience* **27,** 9769–9779 (2007).

234. Ji, D. & Wilson, M. A. Firing rate dynamics in the hippocampus induced by trajectory learning. *Journal of Neuroscience* **28,** 4679–4689 (2008).

235. Dudchenko, P. A. & Wood, E. R. in *Space, time and memory in the hippocampal formation* 253–272 (Springer, 2014).

236. Grieves, R. M., Wood, E. R. & Dudchenko, P. A. Place cells on a maze encode routes rather than destinations. *Elife* **5,** e15986 (2016).

237. Dupret, D., O'neill, J., Pleydell-Bouverie, B. & Csicsvari, J. The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature neuroscience* **13,** 995–1002 (2010).

238. Michon, F., Sun, J.-J., Kim, C. Y., Ciliberti, D. & Kloosterman, F. Post-learning hippocampal replay selectively reinforces spatial memory for highly rewarded locations. *Current Biology* **29,** 1436–1444 (2019).

239. Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D. & Spiers, H. J. Hippocampal place cells construct reward related sequences through unexplored space. *Elife* **4,** e06063 (2015).

240. Wu, C.-T., Haggerty, D., Kemere, C. & Ji, D. Hippocampal awake replay in fear memory retrieval. *Nature neuroscience* **20,** 571–580 (2017).

241. Krause, E. L. & Drugowitsch, J. A large majority of awake hippocampal sharp-wave ripples feature spatial trajectories with momentum. *Neuron* **110,** 722–733 (2022).

242. O'Neill, J., Senior, T. J., Allen, K., Huxter, J. R. & Csicsvari, J. Reactivation of experience-dependent cell assembly patterns in the hippocampus. *Nature neuroscience* **11,** 209–215 (2008).

243. Cheng, S. & Frank, L. M. New experiences enhance coordinated neural activity in the hippocampus. *Neuron* **57,** 303–313 (2008).

244. Eschenko, O., Ramadan, W., Mölle, M., Born, J. & Sara, S. J. Sustained increase in hippocampal sharp-wave ripple activity during slow-wave sleep after learning. *Learning & memory* **15,** 222–228 (2008).

245. Ramadan, W., Eschenko, O. & Sara, S. J. Hippocampal sharp wave/ripples during sleep for consolidation of associative memory. *PloS one* **4,** e6697 (2009).

246. Stella, F., Baracskay, P., O'Neill, J. & Csicsvari, J. Hippocampal reactivation of random trajectories resembling Brownian diffusion. *Neuron* **102,** 450–461 (2019).

247. Siegle, J. H. & Wilson, M. A. Enhancement of encoding and retrieval functions through theta phase-specific manipulation of hippocampus. *elife* **3,** e03061 (2014).

248. Senzai, Y. Function of local circuits in the hippocampal dentate gyrus-CA3 system. *Neuroscience research* **140,** 43–52 (2019).

249. Ylinen, A. *et al.* Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms. *Journal of Neuroscience* **15,** 30–46 (1995).

250. Nakashiba, T., Buhl, D. L., McHugh, T. J. & Tonegawa, S. Hippocampal CA3 output is crucial for ripple-associated reactivation and consolidation of memory. *Neuron* **62,** 781–787 (2009).

251. Davoudi, H. & Foster, D. J. Acute silencing of hippocampal CA3 reveals a dominant role in place field responses. *Nature neuroscience* **22,** 337–342 (2019).

252. Yang, S.-T., Shi, Y., Wang, Q., Peng, J.-Y. & Li, B.-M. Neuronal representation of working memory in the medial prefrontal cortex of rats. *Molecular brain* **7,** 1–13 (2014).

253. Senzai, Y. & Buzsáki, G. Physiological properties and behavioral correlates of hippocampal granule cells and mossy cells. *Neuron* **93,** 691–704 (2017).

254. McNaughton, B. L., Barnes, C. A. & O'Keefe, J. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Experimental brain research* **52,** 41–49 (1983).

255. O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3,** 317–330 (1993).

256. Catanese, J., Viggiano, A., Cerasti, E., Zugaro, M. B. & Wiener, S. I. Retrospectively and prospectively modulated hippocampal place responses are differentially distributed along a common path in a continuous T-maze. *Journal of Neuroscience* **34,** 13163–13169 (2014).

257. Fuller, P. M., Gooley, J. J. & Saper, C. B. Neurobiology of the sleep-wake cycle: sleep architecture, circadian regulation, and regulatory feedback. *Journal of biological rhythms* **21,** 482–493 (2006).

258. Ainge, J. A., Van Der Meer, M. A., Langston, R. F. & Wood, E. R. Exploring the role of context-dependent hippocampal activity in spatial alternation behavior. *Hippocampus* **17,** 988–1002 (2007).

259. Arnol'd, V. I., Dubrovin, B., Kirillov, A., Krichever, I., *et al. Dynamical Systems IV: Symplectic geometry and its applications* (Springer, 2001).

260. Leimkuhler, B. & Reich, S. *Simulating hamiltonian dynamics* (Cambridge university press, 2004).

261. Hairer, E., Lubich, C. & Wanner, G. *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations* (Springer Science & Business Media, 2006).

262. Sandberg, M. & Szepessy, A. Convergence rates of symplectic Pontryagin approximations in optimal control theory. *ESAIM: Mathematical Modelling and Numerical Analysis* **40,** 149–173 (2006).

263. Carlsson, J., Sandberg, M. & Szepessy, A. Symplectic Pontryagin approximations for optimal design. *ESAIM: Mathematical Modelling and Numerical Analysis* **43,** 3–32 (2009).

264. Chyba, M., Hairer, E. & Vilmart, G. The role of symplectic integrators in optimal control. *Optimal control applications and methods* **30,** 367–382 (2009).

265. Peng, H., Gao, Q., Wu, Z. & Zhong, W. Symplectic algorithms with mesh refinement for a hypersensitive optimal control problem. *International Journal of Computer Mathematics* **92,** 2273–2289 (2015).

266. Abe, Y., Nishida, G., Sakamoto, N. & Yamamoto, Y. Symplectic Numerical Approach for Nonlinear Optimal Control of Systems with Inequality Constraints. *International Journal of Modern Nonlinear Theory and Application* **4,** 234 (2015).

267. Wang, X., Peng, H., Zhang, S., Chen, B. & Zhong, W. A symplectic pseudospectral method for nonlinear optimal control problems with inequality constraints. *ISA transactions* **68,** 335–352 (2017).

268. Kalmbach, G. *Quantum measures and spaces* (Springer, 1998).

269. Birkhoff, G. & Von Neumann, J. The logic of quantum mechanics. *Annals of mathematics,* 823–843 (1936).

270. Savage, L. J. *The foundations of statistics* (Courier Corporation, 1972).

271. De Finetti, B. *Theory of probability: a critical introductory treatment* (John Wiley & Sons, 1990).

272. Gotay, M. J. in *Mechanics: from theory to computation* 171–216 (Springer, 2000).

273. Woit, P. *Quantum theory, groups and representations: An introduction* (Springer, 2017).

274. Bohm, D. *Quantum theory* (Dover, 1989).

275. Curtright, T. L., Fairlie, D. B. & Zachos, C. K. *A concise treatise on quantum mechanics in phase space* (World Scientific Publishing Company, 2013).

276. Gutzwiller, M. C. *Chaos in classical and quantum mechanics* (Springer, 2013).

277. Tversky, A. & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive psychology* **5,** 207–232 (1973).

278. Tversky, A. & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *science* **185,** 1124–1131 (1974).

279. Tversky, A. & Kahneman, D. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47,** 263–292 (1979).

280. Tversky, A. & Kahneman, D. The framing of decisions and the psychology of choice. *science* **211,** 453–458 (1981).

281. Tversky, A. & Kahneman, D. Rational choice and the framing of decisions. *Journal of business,* S251–S278 (1986).

282. Allais, P. M. Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica* **21,** 503–546 (1953).

283. Ellsberg, D. Risk, ambiguity, and the Savage axioms. *The quarterly journal of economics,* 643–669 (1961).

284. Schuman, H. & Presser, S. *Questions and answers in attitude surveys: Experiments on question form, wording, and context* (Sage, 1996).

285. Bradburn, N. M., Sudman, S. & Wansink, B. *Asking questions: the definitive guide to questionnaire design–for market research, political polls, and social and health questionnaires* (John Wiley & Sons, 2004).

286. Tversky, A. & Kahneman, D. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review* **90,** 293 (1983).

287. Hampton, J. A. Disjunction of natural concepts. *Memory & Cognition* **16,** 579–591 (1988).

288. Hampton, J. A. Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14,** 12 (1988).

289. Busemeyer, J. R. & Bruza, P. D. *Quantum models of cognition and decision* (Cambridge University Press, 2012).

290. Tversky, A. & Shafir, E. The disjunction effect in choice under uncertainty. *Psychological science* **3,** 305–310 (1992).

291. Conte, E. *et al.* Mental states follow quantum mechanics during perception and cognition of ambiguous figures. *Open Systems & Information Dynamics* **16,** 85–100 (2009).

292. Yukalov, V. I. & Sornette, D. Decision theory with prospect interference and entanglement. *Theory and Decision* **70,** 283–328 (2011).

293. Pothos, E. M. & Busemeyer, J. R. Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences* **36,** 255–274 (2013).

294. Wang, Z., Solloway, T., Shiffrin, R. M. & Busemeyer, J. R. Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences,* 201407756 (2014).

295. Blutner, R. *et al.* Descriptive and Foundational Aspects of Quantum Cognition. *arXiv preprint arXiv:1410.3961* (2014).

296. Khrennikov, A. Y. *Ubiquitous quantum structure* (Springer, 2014).

297. Moreira, C. A. P. *Quantum Probabilistic Graphical Models for Cognition and Decision* PhD thesis (Instituto Superior Técnico, 2017).

298. Wang, Z. & Busemeyer, J. R. A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science* **5,** 689–710 (2013).

299. Fuchs, C. A. & Schack, R. Quantum-bayesian coherence. *Reviews of modern physics* **85,** 1693 (2013).

300. DeBrota, J. B., Fuchs, C. A., Pienaar, J. L. & Stacey, B. C. Born's rule as a quantum extension of Bayesian coherence. *Physical Review A* **104,** 022207 (2021).

301. Deutsch, D. Quantum theory of probability and decisions. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **455,** 3129–3137 (1999).

302. Wallace, D. Quantum probability and decision theory, revisited. *arXiv preprint quant-ph/0211104* (2002).

303. Friederich, S. & Evans, P. W. in *The Stanford Encyclopedia of Philosophy* (ed Zalta, E. N.) Summer 2019 (Metaphysics Research Lab, Stanford University, 2019).