Adaptive learning in humans, brains and neural networks : the role of uncertainty and probabilities

Cédric Foucault

Adaptive learning in humans, brains and neural networks: The role of uncertainty and probabilities

Cédric Foucault

PhD Thesis

Carried out from September 2020 to December 2023

Directeur de thèse : Florent Meyniel

Cognitive Neuroimaging Unit, NeuroSpin

Defended on 18 December 2023

Before a jury composed of:

Nils Kolling Université Lyon 1	Rapporteur
Angela Yu Technical University of Darmstadt	Rapporteuse
Christopher Summerfield University of Oxford	Examinateur
Stefano Palminteri ENS - Université PSL	Examinateur et Président du jury
Florent Meyniel Université Paris-Saclay	Directeur de thèse

Abstract | *Résumé*

When learning, the brain faces a difficult problem. It is learning from an environment that is both stochastic and dynamic, i.e. that is characterised by probability distributions that are subject to change. Stochasticity encourages the brain to integrate a large number of past observations in order to achieve greater precision in learned knowledge, while the propensity to change encourages the brain to consider only the most recent observations. This tension leads to a tradeoff in the learning rate to be used, which the brain should ideally adapt constantly. How does the brain achieve such adaptive learning? According to normative theory, it should constantly represent the uncertainty of its knowledge through probability distributions and update its knowledge by weighting the different probabilities, as described by the computations of dynamic Bayesian inference (also known as Bayesian filtering). Guided by this theory, my thesis work has investigated the computational and neural bases of adaptive learning in humans through three studies.

In the first study (Foucault & Meyniel, submitted), I examined human behaviour during magnitude and probability learning and tested its alignment with normative theory. Using a new experimental paradigm, I measured the dynamics of human learning rates observation by observation, and related them to the computational factors that govern the adaptations of the learning rate according to normative theory. I have shown that humans adapt their learning rate dynamically and normatively, guided by the uncertainty, which plays a dominant role in probability learning. Overall, I found that human behaviour closely followed normative theory, providing a strong empirical ground for my overarching thesis.

How could the normative adaptive learning observed in humans be realised in the brain at the computational and mechanistic level? In the second, theoretical study (Foucault & Meyniel, 2021, published in eLife), I explored this question by studying the computational mechanisms that enable neural networks to achieve normative adaptive learning capabilities. I have shown that with a simple gating mechanism, which could be implemented in the brain by the locus coeruleus-norepinephrine system, a tiny recurrent neural network can perform quasi-optimal Bayesian learning.

This mechanism is key to the network's ability to evaluate and use uncertainty to adapt its learning rate, as found by decoding and manipulating the uncertainty represented in the network's internal recurrent dynamic activities, and it is also key to other adaptive capabilities.

A fundamental quantity in our theory is probability. As shown in the first study, humans are clearly capable of estimating probability, but how is it represented in the brain? Neurally, the representation of probability has remained elusive. My third study (Foucault*, Bounmy* et al., in prep.), attempts to answer this question by using a new method that combines encoding models with approximation theory to model the neural coding of probability without making assumptions about the form of the code. This method revealed a previously unknown neural code for probability in human brain activity measured by fMRI during learning within the dorsolateral prefrontal and intraparietal cortices. I characterised this neural code by reconstructing its tuning curves. I found that, compared to the neural code for uncertainty (which had already been observed in previous studies and which I further characterised in this study), the neural code for probability was highly nonlinear and non-monotonic.

Altogether, my thesis provides theoretical and empirical insights into the computational and neural bases of the human learning process and the role of uncertainty and probabilities in this process.

Résumé (version française de l'Abstract)

Lors de l'apprentissage, le cerveau est confronté à un problème difficile. Il apprend d'un environnement qui est à la fois stochastique et dynamique, c'est-à-dire qui se caractérise par des distributions de probabilité qui sont susceptibles de changer. La stochasticité incite le cerveau à intégrer un grand nombre d'observations passées pour arriver à une meilleure précision des connaissances acquises, tandis que la propension au changement l'incite à ne prendre en compte que les observations les plus récentes. Cette tension donne lieu à un compromis dans le taux d'apprentissage à utiliser, que le cerveau devrait idéalement adapter en permanence. Comment le cerveau parvient-il à un tel apprentissage adaptatif ? Selon la théorie normative, il devrait constamment représenter l'incertitude de ses connaissances par des distributions de probabilité, et les mettre à jour en pondérant les différentes probabilités, comme décrit dans les calculs de l'inférence bayésienne dynamique (aussi appelée filtrage bayésien). Guidés par cette théorie, mes travaux de thèse ont examiné les bases computationnelles et neurales de l'apprentissage adaptatif chez les humains à travers trois études.

Dans la première étude (Foucault & Meyniel, soumise pour publication), j'ai examiné le comportement humain lors de l'apprentissage des magnitudes et des probabilités et testé son alignement avec la théorie normative. Grâce à un nouveau paradigme expérimental, j'ai mesuré les dynamiques du taux d'apprentissage des humains observation après observation, et les ai mises en relation avec les facteurs computationnels qui régissent les adaptations du taux d'apprentissage selon la théorie normative. J'ai montré que les humains adaptent leur taux d'apprentissage de façon dynamique et normative, guidés par l'incertitude, qui joue un rôle prépondérant dans l'apprentissage des probabilités. Globalement, j'ai constaté que le comportement humain suivait de près la théorie normative, offrant ainsi un fondement empirique solide pour ma thèse générale.

Comment l'apprentissage adaptatif normatif observé chez les humains peut-il être réalisé dans le cerveau au niveau computationnel et mécanistique ? Dans la deuxième étude, théorique (Foucault & Meyniel, 2021, publiée dans eLife), j'ai

exploré cette question en étudiant les mécanismes computationnels qui permettent à des réseaux de neurones de réaliser des capacités d'apprentissage adaptatif normatives. J'ai montré qu'avec un simple mécanisme de gating, qui pourrait être implémenté dans le cerveau par le système locus coeruleus-norepinephrine, un très petit réseau de neurones récurrent peut réaliser un apprentissage bayésien quasi-optimal. Ce mécanisme est essentiel pour la capacité du réseau à évaluer et utiliser l'incertitude pour adapter son taux d'apprentissage, comme je l'ai montré en décodant et en manipulant l'incertitude représentée dans les activités dynamiques récurrentes internes au réseau, et il est aussi essentiel pour d'autres capacités adaptatives.

Une quantité fondamentale de notre théorie est la probabilité. Comme l'a montré la première étude, les humains sont clairement capables d'estimer la probabilité, mais comment est-elle représentée dans le cerveau ? Au niveau neural, la représentation de la probabilité est restée insaisissable. Ma troisième étude (Foucault*, Bounmy* et al., en préparation) tente de répondre à cette question en utilisant une nouvelle méthode alliant des modèles d'encodage à la théorie de l'approximation afin de modéliser le codage neural de la probabilité sans faire de supposition sur la forme du code. Cette méthode a révélé un code neural de la probabilité jusqu'alors inconnu dans l'activité cérébrale humaine mesurée par IRMf lors de l'apprentissage dans les cortex préfrontal dorsolatéral et intrapariétal. J'ai caractérisé ce code neural en reconstruisant ses *tuning curves* (fonctions d'encodage). J'ai ainsi constaté que, par rapport au code neural de l'incertitude (qui avait déjà été observé dans de précédentes études et que j'ai caractérisé plus en détail dans cette étude), le code neural de la probabilité était hautement non-linéaire et non-monotone.

Dans l'ensemble, ma thèse apporte un éclairage théorique et empirique sur les bases computationnelles et neurales du processus d'apprentissage humain et le rôle de l'incertitude et des probabilités dans ce processus.

Remerciements | Acknowledgements

J’ai eu le privilège de faire ma thèse sous la direction de Florent Meyniel. Je souhaite le remercier pour le rôle capital qu’il a joué dans le développement de cette thèse et dans mon développement en tant que scientifique, qui mérite d’être détaillé.

En tant que directeur de thèse et collaborateur scientifique, Florent m’a suivi et fourni des retours réguliers au cours de réunions individuelles hebdomadaires, de discussions en ligne et par email, qui ont alimenté et guidé mes recherches. Il a joué un rôle primordial dans la qualité scientifique de notre travail grâce à de nombreuses contributions intellectuelles. Je le remercie pour ses relectures attentives et détaillées, et de m’avoir aiguillé et fait profiter de son expérience. Je le remercie également de m’avoir laissé de l’autonomie et de l’indépendance dans le travail, de m’avoir fait progressivement confiance et de manière raisonnée, et de m’avoir permis de co-encadrer un étudiant avec lui.

Je suis aussi reconnaissant à Florent d’avoir créé une équipe, l’équipe Computational Brain, dont j’ai eu le chance de faire partie. En tant que chef d’équipe, il a su créer les conditions pour que nous soyons une équipe soudée et unie, grâce notamment à un chevauchement fructueux dans nos sujets de recherche, nos méthodes, nos outils de travail et à nos pratiques communes.

Enfin, je souhaite remercier Florent en tant que scientifique en général, parce qu’il est un exemple pour moi à de nombreux égards. Il incarne des valeurs qui me sont chères telles que l’intégrité, la méticulosité, l’attention au détail, la finesse, la recherche de qualité, l’humilité et la capacité à changer d’avis. J’ai beaucoup appris de son calme, de sa capacité à être nuancé, et de son vocabulaire riche et bien employé.

Je remercie les autres investigateurs principaux qui m’ont fourni du soutien et des conseils lors de ma thèse : Evelyn Eger, Bertrand Thirion, Christophe Pallier, Sophie Herbst, Jean-Rémi King, Mehdi Khamassi. Ils servent également d’exemple pour moi.

Je remercie les membres de mon jury de thèse : Nils Kolling, Angela Yu, Chris Summerfield, et Stefano Palminteri. Je vous suis reconnaissant d'avoir accepté de me lire, d'évaluer mon travail et de participer à ma soutenance, ce qui représente un investissement de temps important.

Je souhaite témoigner ma profonde gratitude aux professeurs des universités qui m'ont encadré et ont joué un rôle clé dans mon développement en amont de la thèse : Claire Sergent, Thérèse Collins, Jérôme Sackur. J'en profite pour remercier plus largement le Cogmaster, sans lequel je n'aurais pas pu rejoindre le domaine des neurosciences cognitives et computationnelles.

Je remercie NeuroSpin, mon laboratoire, et l'hôpital Sainte-Anne et l'Institut de Neuromodulation, de m'avoir fourni un lieu de travail et un environnement propice à faire du bon travail.

Je remercie l'ENS Paris-Saclay, de m'avoir accordé un contrat doctoral spécifique normalien pour financer ma thèse, et de m'avoir accordé ainsi leur confiance bien que j'ai exercé dans le privé auparavant entre la sortie de l'école et le début de la thèse.

Je suis reconnaissant envers les organisateurs des conférences Cognitive Computational Neuroscience (CCN) 2022 et 2023. Ces conférences ont été des occasions uniques pendant ma thèse de me connecter à une communauté scientifique et aux membres de cette communauté.

Je ne remercierai jamais assez ma mère et mon père, Joëlle Bidault et Alain Foucault. Je vous remercie en particulier de m'avoir aidé à rentrer en France en vue de réaliser cette thèse, après avoir passé plusieurs années en Californie, et pendant ma thèse, de m'avoir écouté et apporté du soutien.

Enfin, je remercie chaleureusement mes collègues de l'équipe Computational Brain que j'ai côtoyés pendant plusieurs années et qui sont devenus des amis proches : Maëva L'Hôtellier, Alexander Paunov, Tiffany Bounmy, Caroline Bévalot, Audrey Mazancieux. La vie de thésard aurait été beaucoup plus morose et beaucoup moins riche sans nos interactions quotidiennes. J'ai adoré travailler avec vous et je me réjouis de toutes les prochaines occasions que nous aurons d'interagir.

Acknowledgments (English version of the *Remerciements*)

I had the privilege of doing my thesis under the supervision of Florent Meyniel. I would like to thank him for the crucial role he played in the development of this thesis and in my growth as a scientist, which deserves to be detailed.

As a thesis supervisor and scientific collaborator, Florent monitored my progress and provided regular feedback during weekly individual meetings, online discussions, and emails, which fueled and guided my research. He played a crucial role in the scientific quality of our work through numerous intellectual contributions. I thank him for his careful and detailed reviews, and for guiding me and sharing his experience with me. I also thank him for giving me autonomy and independence in my work, for trusting me in a gradual and reasoned manner, and for allowing me to co-supervise a student with him.

I am also grateful to Florent for having created a team, the Computational Brain team, which I was lucky enough to be part of. As team leader, he created the conditions for us to be a cohesive and united team, thanks in particular to fruitful overlaps in our research topics, methods, work tools, and to common practices.

Finally, I would like to thank Florent as a scientist in general, because he is an example to me in many respects. He embodies values that are dear to me such as integrity, meticulousness, attention to detail, finesse, pursuit of quality, humility, and the ability to change one's mind. I have learned a lot from his calmness, his ability to be nuanced, and his rich and well-chosen vocabulary.

I thank very much the other principal investigators who provided me with support and advice during my thesis: Evelyn Eger, Bertrand Thirion, Christophe Pallier, Sophie Herbst, Jean-Rémi King, Mehdi Khamassi. They also serve as examples to me.

I thank the members of my thesis jury: Nils Kolling, Angela Yu, Chris Summerfield, and Stefano Palminteri. I am grateful to you for agreeing to read my work, evaluate it, and participate in my defence, which represents a significant time investment.

I would like to express my deep gratitude to the professors who supervised me and played a key role in my development prior to the thesis: Claire Sergent, Thérèse Collins, Jérôme Sackur. I would also like to thank the Cogmaster more broadly, without which I would not have been able to join the field of cognitive computational neuroscience.

I thank NeuroSpin, my laboratory, and Sainte-Anne Hospital and the Neuromodulation Institute for providing me with a workplace and an environment conducive to doing great work.

I thank ENS Paris-Saclay for granting me a doctoral contract (contrat doctoral spécifique normalien) to fund my thesis, and for trusting me even though I had worked in the private sector before, between graduating from the school and starting my thesis.

I am grateful to the organisers of the Cognitive Computational Neuroscience (CCN) 2022 and 2023 conferences. These conferences were unique opportunities during my thesis to connect with a scientific community and its members.

I cannot thank my mother and father enough, Joëlle Bidault and Alain Foucault. I thank you especially for helping me return to France in view of pursuing this thesis, after I had spent several years in California, and during my thesis, for listening to me and supporting me.

Finally, I would like to give my warm thanks to my colleagues from the Computational Brain team whom I have worked with for several years and who have become close friends: Maëva L'Hôtellier, Alexander Paunov, Tiffany Bounmy, Caroline Bévalot, Audrey Mazancieux. My PhD life would have been much duller and much less enriching without our daily interactions. I've loved working with you and I look forward to all the future opportunities we will have to interact.

Table of contents

Chapter I: General introduction

1. Adaptive learning in stochastic and dynamic environments

Today, Emily came into the room and sat down at her desk without saying a word. She usually says hello, but not always. This time, though, there was something more. I think it's the way she walked: it was slower than usual. It got me wondering. I started to wonder if there might be something wrong. Then, during the break, she remained very quiet. But she gave me a big smile when I held the door for her, so I thought maybe there was nothing to worry about. Except that later on, Alex's children came by and she seemed completely disinterested, even though she usually loves playing with them. That's when I became really concerned. I had to ask her.

> "Is everything okay, Emily?"
>
> — A made-up short story.

1.1. Learning in a stochastic and dynamic environment

In the short story above, the narrator makes observations about their colleague's behaviour and tries to learn about her state. This state is not directly observable to the narrator: it is a hidden state. Based on the observations, the narrator can only estimate this state.

Estimating the person's state is not easy because the relationship between the observations the narrator can make and the underlying state of the person is uncertain. For example, the fact that the person does not say hello or that she walks slower than usual is not necessarily indicative of a bad state (the person could simply be thinking about something else, among others). Similarly, even though it is generally a positive sign, the fact that the person smiles does not necessarily mean that she is doing well. The observation of any given behaviour is only more or less probable, but not certain, under one state or another. This element of probability, the absence of certainty in the correspondence between observations and the hidden state, is called ***stochasticity***.

The person's state is also ***dynamic***. It can change from one day to another (one day the person is fine, the next day she is not) or even from one moment to another. For this reason, the observer must constantly re-evaluate their estimate of the person's

state. This process by which an observer updates their knowledge as observations are made is referred to as a ***learning*** process (formally conceptualised as a sequential inference process) in this thesis.

Learning a hidden state in a stochastic and dynamic environment is a complex and ubiquitous task. As illustrated in the situation described above, in social interactions, we are constantly learning about the mental states, physical states, intentions, interests, opinions of others, and other dynamic characteristics of others which we can only have uncertain knowledge about based on their behaviours. In medical monitoring, the patient or the doctor following the patient must constantly watch for certain symptoms that are more or less indicative of an abnormal state in order to successfully detect the onset of an epileptic seizure or a psychotic break, for example. More mundane examples include our daily commute: we learn our average travel time in the presence of stochasticity (travel times for the same route typically vary randomly from one day to another) and unpredictable changes (average travel time can be delayed due to roadworks, for example).

To perform such learning, the brain must evaluate probabilities. As observations are received, it must update its knowledge, taking into account the probabilities that such observations occur according to different possible states, and taking into account the probabilities it can assign to those states based on prior knowledge or learning. The probabilistic nature of learning serves as the basis for the theoretical framework used throughout my thesis: probability theory.

Before further describing learning in a probabilistic framework, I provide some historical elements on the study of learning in psychology and neuroscience that have led the field to adopt this framework.

1.2. Learning in a probabilistic inference framework: A brief historical context

The study of learning in animals began with associative learning and conditioning experiments (Dickinson, 1980; Mackintosh, 1983; Skinner, 1938). These experiments investigate how animals learn to predict events significant for their survival and well-being, such as the receipt of food or a painful stimulus, by association with a preceding event. In the case of classical conditioning (or Pavlovian conditioning), the

predictive event is the occurrence of a stimulus (initially neutral, called "conditioned stimulus"), such as a bell sound or a light. In the case of operant conditioning (or instrumental conditioning, Skinnerian conditioning), the predictive event is an action performed by the animal, such as pressing a lever. When the experiment begins, the predictive and predicted events are paired, and the animal learns through repetitions that the first event predicts (deterministically or probabilistically, depending on the experiment) the occurrence of the second event. Learning is measured in these experiments by the increase in the frequency of a certain behaviour of the animal. In Pavlov's experiments, it was the dogs' salivation in response to the sound of the bell after the bell had been associated with the receipt of food. In Skinner's experiments, it was the increase in lever presses (by rats or pigeons) after this action had been associated with the receipt of food.

Originally, the interpretation of conditioning experiments, in accordance with the dominant behaviourist viewpoint of the time (Watson, 1913), was that learning could be reduced to the formation of a direct association between the stimulus (such as the sound of a bell) and the behavioural response (such as salivation). With the emergence of cognitivism, it became clear that this restrictive interpretation was difficult to maintain. It was more useful to postulate that the animal could form an *internal representation* of the associations between different events, and that the change in behaviour (such as salivation) only indexed this representation that had been formed (Dickinson & Mackintosh, 1978). Such internal representations could account for many phenomena that were difficult to explain from a purely behaviourist standpoint. One of the most famous example was explaining how rats could quickly navigate to the exit of a maze to obtain food on their first attempt, without this behaviour having been reinforced beforehand, as long as the rats had had the opportunity to freely move in the maze on previous days (during which they did not try to exit the maze). According to Tolman, the rats had formed an internal representation of the maze (a "cognitive map") that then allowed them to know how to navigate to a specific location (Tolman, 1948). In the conditioning experiments themselves, one example of a phenomenon unaccounted for by behaviourism was that animals are sensitive to the time interval between stimuli: dogs do not immediately salivate at the sound of the bell, and if the time interval between the bell and the food is increased, they start salivating later and later after the bell (Gallistel &

Gibbon, 2000). This can be explained assuming that dogs learn and maintain a representation of the time interval between stimuli. More generally, according to a cognitive view of learning, animals possess and update internal representations that guide their predictions. Such internal representations can be, for example, a representation of the reliability with which an event predicts the occurrence of another event, or a representation of the future rewards that are expected to follow a given observation.

Theories on how internal representations are learned have long focused on basic learning rules. One of the most influential is the delta-rule, made famous by the Rescorla-Wagner model, and involved in many other models, such as temporal difference learning and various kinds of reinforcement learning models (Bush & Mosteller, 1951; Rescorla & Wagner, 1972; Sutton & Barto, 2018; Widrow & Hoff, 1960). According to this rule, following an observation, which has a given value, the internal representation is updated in proportion to the prediction error, which is the difference between the observed value and the predicted value. This is expressed by the equation $\Delta v(t) = \alpha\, \delta(t)$, where $\Delta v(t)$ is the amount of update in the representation, $\delta(t)$ is the prediction error, and α is the constant of proportionality. Despite the many successes they have allowed, these basic learning rules have limitations. How can we explain, for example, with a constant α, that learning is sometimes very slow (e.g. in conditioning experiments, many trials can be required for the animal to learn an arbitrary association) and sometimes very fast (e.g. only a single intoxication is needed to learn to avoid a certain food, taste, or smell)? Even within a domain, the speed of learning can vary over time. For example, in language learning, the speed of learning can go from 2 to 3 words per week to 10 words per day in learning to associate a word with its referent (Bloom, 2013).

One approach to explaining new phenomena would be to resort to more sophisticated learning rules, as Pearce and Hall did, for example (Pearce & Hall, 1980). However, a principles-based approach may be more enlightening. One such approach is that based on the principles of ***probabilistic inference***, which I unpack here. First, learning is treated as an *inference* process within an internal model of the world. Internal representations correspond to variables of that model, and the model specifies how the variables are related to each other and related to the observations

that the learner can make (see e.g. **Figure 1**). Learning consists of inferring what the values of these variables could be given the observations that have been made. Modelling the learning process then amounts to formally specifying the internal model, and solving the inference problem within that model. Second, the inferential perspective on learning is also combined with a *probabilistic* (or statistical) perspective. The probabilistic perspective is motivated by the fact that learning situations in the real world often involve incomplete information or probabilistic outcomes, which make the inference process itself probabilistic (Chater et al., 2006).

Following this approach, learning has been increasingly considered in terms of probabilistic inference (Courville et al., 2004, 2006; Dayan et al., 2000; Dayan & Long, 1997; Gershman et al., 2010). In the words of Peter Dayan, *"any system, natural or artificial, that estimates the current predictive relationship between stimuli and reward based on past observations can be described as making statistical inferences"* (Dayan et al., 2000). This trend has benefited from concomitant progress in the mathematics and computer science of probabilistic models. These advances have provided the theoretical and practical tools to define and compute the optimal solution to the learning problem, once specified in terms of probabilistic inference within a model. The optimal solution is a ***normative*** model, that is, a model that explains how the problem *should* be solved, according to an optimality principle. The normative probabilistic framework thus offers an explanation of both how and why subjects learn as they do, and allows researchers to identify quantities and principles of the probabilistic theory that may underlie patterns of experimental data and are likely to generalise to other situations (Courville et al., 2006; Dayan et al., 2000).

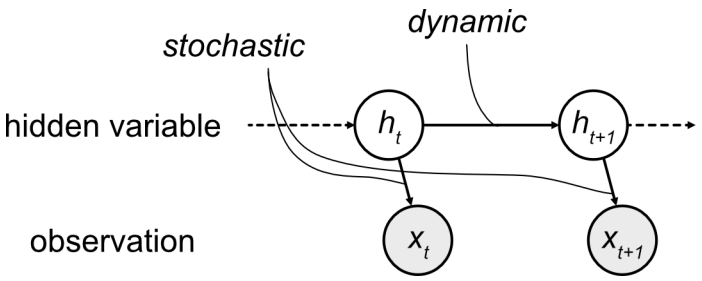


Figure 1. Graphical model of the stochastic and dynamic environment in which learning is taking place. The learner tries to infer the current value of the hidden variable (h_t) based on past observations ($x_{1:t}$). The arrows represent conditional dependencies in the environment. The observation is generated depending on the hidden variable according to a conditional probability distribution $p(x_t \mid h_t)$ which describes stochasticity, and the next value of the hidden variable is generated depending on the previous one according to a conditional probability distribution $p(h_{t+1} \mid h_t)$ which describes the dynamics of the environment.

1.3. Essential concepts for the study of adaptive learning

Several concepts have emerged in the previous sections. In this section, I will clarify these concepts with precise definitions and link them together by listing them in an organised manner. As I do so, I will introduce a formalism with notations (summarised in **Box 1**) that will be used throughout the document. This formalism will serve as a unifying framework to explain a large number of tasks and learning models which are often treated separately in the literature, and whose connections will be made explicit in my thesis thanks to this formalism.

The word "learning" has a different meaning depending on the context in which it is used, but in a broad sense it usually involves an acquisition of knowledge (Greeno, 1980). In the context of this thesis, *learning* is defined as *the process by which knowledge is updated*. The learner's *knowledge*, which will also be referred to interchangeably as *"belief"*, will concern one or more variables of the environment, which are hidden from the learner (i.e. not directly observable by them). These environment variables are interchangeably referred to as *hidden variable, hidden quantity*, or *hidden state*, and denoted by the symbol *h* (**Figure 1)**. For simplicity, I will generally use the singular "hidden variable" whether the environment variables are

Box 1. Symbols and notation.	
t	Time step
h_t	Hidden variable at time step t (scalar or vector composed of several scalar variables)
x_t	Observation at time step t
$x_{1:t}$	Sequence of observations from the first time step to time step t
v_t	Estimate following the observation at time step t
u_t	Uncertainty prior to the observation at time step t
α_t	(Apparent) Learning rate following the observation at time step t
$p(z)$	Probability density of continuous variable z or probability of discrete variable z
$p(z\|y)$	Conditional probability density or conditional probability of z given y
$\int f(z)dz$	Integral over the entire domain of z

composed of one or several scalar variables, treating the hidden variable as a vector composed of the different scalar variables when there are several. I will generally prefer the term “quantity” over “state” because in the contexts I am studying the most, the hidden variable has continuous rather than categorical values. In experiments, the environment and its hidden variable are defined and manipulated by the experimenter.

During the learning process, the learner is learning the value of the hidden variable based on the *observations* received *sequentially* from the environment (**Figure 1**). Each observation may trigger an update in the learner’s knowledge. The nature of the observation depends on the learning context. It can be, for example, the occurrence of a reward. The observation will also be referred to interchangeably as *“outcome”* (the term “outcome” has two senses: the realisation of a random variable, or the result of an action; it turns out that in the learning contexts considered in this thesis, the observations are always realisations of a random variable, and can sometimes be the result of an action). In experiments, the observations and their relationship with the hidden variable are also under the experimenter’s control. The observation will be denoted by the symbol x.

As each observation can trigger an update, it defines a discrete unit of time for the learning process: each observation defines one *time step*. In the context of

experiments, a time step can also be referred to as a *trial*. The time step will be denoted by the symbol *t*. It will be used as a subscript to index each value in a sequence: x_t is the value of the *t*-th observation of the sequence, and h_t is the value of the hidden variable that conditioned the *t*-th observation. The subsequence comprising all values from the start of the sequence until the *t*-th observation is denoted *1:t*, that is, $x_{1:t} = (x_1, \ldots, x_t)$.

All the learning contexts considered here require the learner to estimate the value of the hidden variable based on their acquired knowledge. The learner's *estimate* (which is a point estimate of a scalar hidden variable) will be denoted by the symbol *v*.

The learner may associate a degree of *uncertainty* to their estimate. This type of uncertainty is more completely referred to as *"estimation uncertainty"* or *"belief uncertainty"* (see **Note 1**). Uncertainty will be denoted by the symbol *u*.

Both the estimate *v* and the associated uncertainty *u* are derived based on the knowledge acquired and updated by the learner during the learning process. Knowledge is conceived of as an internal representation of the learner that researchers are trying to elucidate. In the probabilistic framework, it formally corresponds to a probability distribution over the hidden variable. This will be explained in quantitative, computational terms in section 3.

A key descriptive tool, used in this thesis to characterise the learning process and define adaptive learning, is the *instantaneous*, *apparent learning rate*, which will simply be referred to as ***learning rate***. The learning rate will be denoted by the symbol *α*. The learning rate quantifies the apparent weight assigned by the learner to a given observation in the updating of their estimate. It is calculated as the ratio between the change in estimate elicited by the observation to the difference between observation and the prior estimate, as described by equation **[1]** below:

$$\alpha_t = (v_t - v_{t-1}) / (x\_t - v\_t\text{-}1) \quad \textbf{[1]}$$

This measure of learning rate is *instantaneous* because it is calculated with respect to a given time step (i.e. observation) in the sequence. It is *apparent* because, unlike the delta-rule introduced previously and other models where the learning rate is also

a true parameter of the learning process, this learning rate measure is a *descriptive* measure that makes no assumptions about the underlying learning process.

Note 1: Definition of uncertainty.

In this thesis, the term "uncertainty" is specifically referring to the uncertainty associated with the estimates, which is a form of epistemic uncertainty. This type of uncertainty is known as ***estimation uncertainty*** or ***belief uncertainty***. There are other types of uncertainty, such as *decision uncertainty* (uncertainty about one's decision being correct), and *outcome uncertainty* (about an upcoming outcome, also known as *unpredictability*, or *risk* in the context of risky choice). A discussion of the different types of uncertainty is out of the scope of this thesis (see e.g. (Bach & Dolan, 2012) for a review discussing different types of variables about which one can be uncertain, and (Walker et al., 2023) for a review of studies investigating different types of uncertainty). Here, I only want to highlight that these different types of uncertainty have distinct functional roles, and may have distinct neural bases. This is why I am careful to be specific about which type of uncertainty I am referring to and not to mix it together with other types of uncertainty.

The term estimation *confidence* or *precision* is sometimes used instead of *uncertainty* in the articles of my thesis and the articles I am referring to in the references. In those articles, confidence, precision and uncertainty are all referring to the same type of uncertainty. According to the mathematical definition I am using in this thesis for uncertainty (equation **[6]** in section 3), and the mathematical definition of confidence often used, *confidence = ½ log(precision) = -log(uncertainty)* (Meyniel et al., 2015).

The estimate, the learning rate, and the uncertainty will be indexed with respect to the observation (x_t) received at time step t in this way: v_t is the estimate *after* observing x_t, α_t is the learning rate for the update from v_{t-1} to v_t, and u_t is the uncertainty about the estimate *before* observing x_t. I chose this indexing because a normative hypothesis investigated in this thesis is that the uncertainty of the estimate modulates the subsequent update of that estimate, i.e. that u_t modulates α_t (this will

be covered in section 3.3). This indexing aligns in time the modulating variable (u_t) and the modulated variable (α_t).

1.4. Learning contexts involving stochastic and dynamic environments

In this section, we will review the paradigms used to study learning in stochastic and dynamic environments, and we will see that they can be grouped into several learning contexts that will be characterised using three axes. This characterization will be used throughout the document and will illuminate the understanding of the relationships between different studies and the challenges in learning that arise specifically in certain contexts.

All the learning contexts studied here involve stochastic and dynamic environments (**Figure 1**), two characteristics that, as illustrated at the beginning of the introduction, are ubiquitous in real life. Let's review them using the concepts defined in the previous section.

> ***Stochastic***: This means that the relationship between the hidden variable h_t and the observation x_t is not deterministic but probabilistic. The degree of stochasticity corresponds to the degree of variability of x_t given h_t according to the probability distribution of the environment.
>
> ***Dynamic***: This means that the hidden variable is subject to change from one time step to another, that is, that h_{t+1} can be different from h_t. Moreover, these changes have a degree of variability (in the time at which they occur, and/or their amplitude) that is also most often probabilistic.

Numerous experimental paradigms have been created to study learning in stochastic and dynamic environments. I have made an inventory of these paradigms and listed representative studies for each of these paradigms in **Table 1**. Examples are shown in **Figure 2**, **3**, and **4**.

Table 1: Representative studies of different learning contexts.

Study	Type of learning	Type of dynamics	Type of response
Daw et al. (2006)	Magnitude learning	Random walk	Choice
Jepma et al. (2011)	Magnitude learning	Random walk	Choice
Speekenbrink et al. (2015)	Magnitude learning	Random walk	Choice
Findling et al. (2019)	Magnitude learning	Random walk	Choice
Fan et al. (2023)	Magnitude learning	Random walk	Choice
Lee et al. (2020)	Magnitude learning	Random walk	Estimate
Ossmy et al. (2013)	Magnitude learning	Change points	Choice
Glaze et al. (2015)	Magnitude learning	Change points	Choice
Murphy et al. (2021)	Magnitude learning	Change points	Choice
Nassar et al. (2010)	Magnitude learning	Change points	Estimate
Nassar et al. (2012)	Magnitude learning	Change points	Estimate
McGuire et al. (2014)	Magnitude learning	Change points	Estimate
Jepma et al. (2016)	Magnitude learning	Change points	Estimate
Vaghi et al. (2017)	Magnitude learning	Change points	Estimate
Prat-Carrabin et al. (2021)	Magnitude learning	Change points	Estimate
Gershman et al. (2009)	Probability learning	Random walk	Choice
Walton et al. (2010)	Probability learning	Random walk	Choice
Fouragnan et al. (2019)	Probability learning	Random walk	Choice
Behrens et al. (2007)	Probability learning	Change points	Choice
Browning et al. (2015)	Probability learning	Change points	Choice
Pulcu et al. (2017)	Probability learning	Change points	Choice
Cook et al. (2019)	Probability learning	Change points	Choice
Gallistel et al. (2014)	Probability learning	Change points	Estimate
Meyniel et al. (2015)	Probability learning	Change points	Estimate
Heilbron et al. (2019)	Probability learning	Change points	Estimate

I have categorised the paradigms along three axes: the ***type of learning*** (*magnitude learning* or *probability learning*), the ***type of dynamics*** (*random walk dynamics* or *change point dynamics*), and the ***type of response*** required from the subject (*choice* or *estimate*). This categorisation defines different learning contexts and highlights the commonalities and differences between contexts, which have implications for the learning process as we will see. Each axis of the categorisation is detailed below.

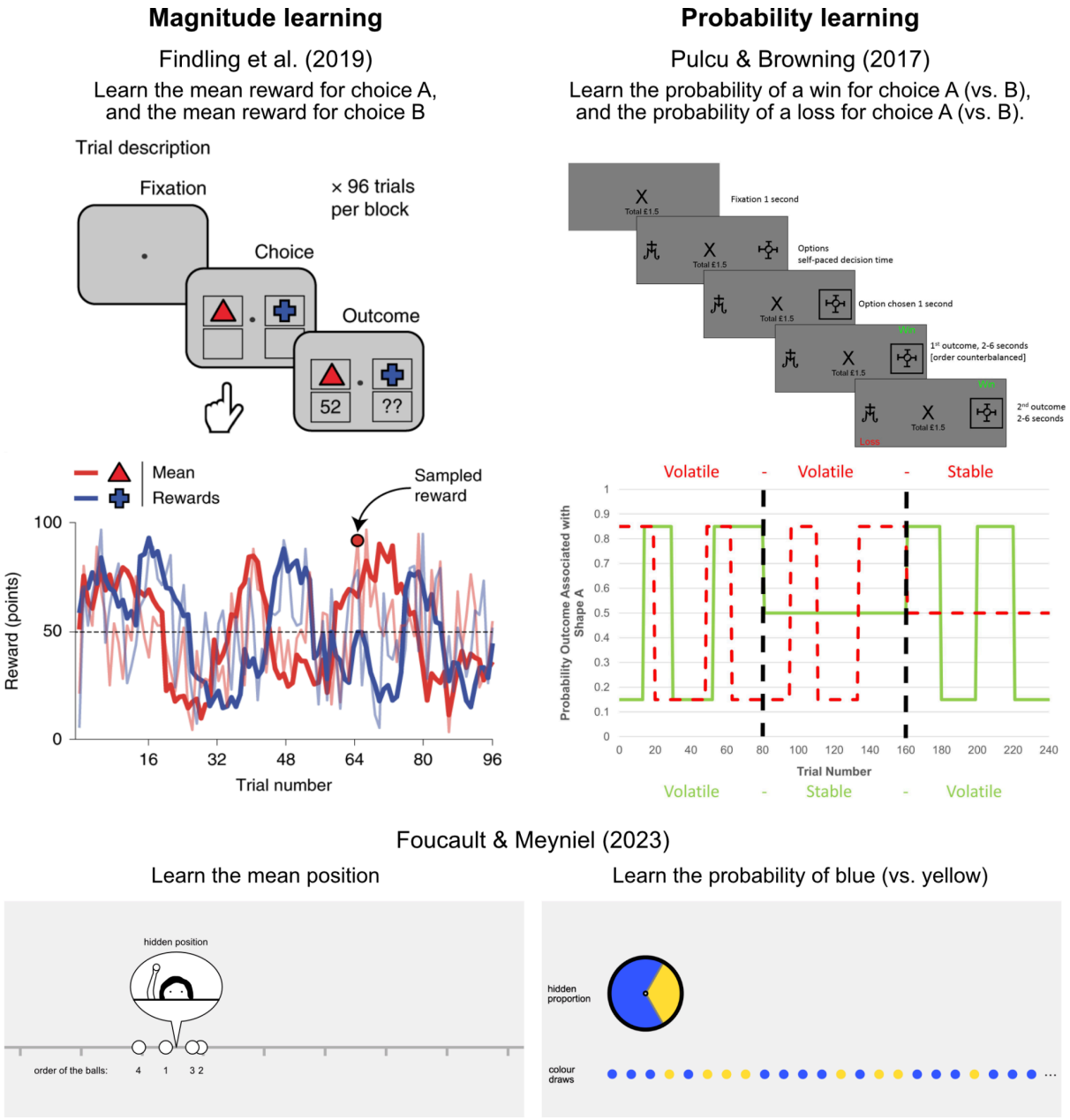


Figure 2. Types of learning. Panels adapted from (Findling et al., 2019; Foucault & Meyniel, 2023; Pulcu & Browning, 2017).

Two types of learning can be involved: ***magnitude learning***, when the hidden quantities being learned (*h*) are magnitudes, or ***probability learning***, when they are probabilities (**Figure 2**). Due to their pervasiveness in real life, magnitudes and probabilities have a central place in psychology and neuroscience. Magnitudes and probabilities often encountered in experiments include the magnitude and probability of occurrence of a reward (e.g. food or money) in decision-making studies (Behrens et al., 2007; Doya, 2008; Edwards, 1954; Fiorillo, 2003; Kahneman & Tversky, 1979;

Skinner, 1953), the sensory magnitude (e.g. lightness, loudness, size, position, angle, motion, number) and probability of occurrence of a stimulus in perception studies (Shepard, 1987; Stevens, 1957; Summerfield & de Lange, 2014; Tudusciuc & Nieder, 2007), and the salience of an attentional cue or the probability that it correctly predicts a target in attention studies (Posner et al., 1980; Treisman & Gelade, 1980; Yu & Dayan, 2005). In learning studies (the focus of this thesis) where magnitudes or probabilities are learned, the magnitudes to be learned are typically the average size of rewards or punishments (Daw et al., 2006; Findling et al., 2019; Jepma & Nieuwenhuis, 2011; Speekenbrink & Konstantinidis, 2015), or the average position (Jepma et al., 2016; Lee et al., 2020; McGuire et al., 2014; Prat-Carrabin et al., 2021), angular direction (J. X. O'Reilly et al., 2013; Vaghi et al., 2017), or symbolic number (Nassar et al., 2010, 2012) of the presented stimuli. The learned probabilities are typically the probabilities of occurrence of rewards or punishments, or of neutral stimuli perceived in the visual, auditory, or nociceptive modalities (Behrens et al., 2007; Browning et al., 2015; Fouragnan et al., 2019; Gallistel et al., 2014; Gershman et al., 2009; Meyniel et al., 2015; Pulcu & Browning, 2017; Walton et al., 2010).

One difference between magnitude learning and probability learning is that the hidden variable belongs to the same domain as the observations in magnitude learning (since the hidden variable is typically the average magnitude, i.e. the mean of the generative distribution of the observation), but does not belong to the same domain as the observations in probability learning (since the hidden variable is typically the probability that a certain category of observation occurs). This may induce differences in the learning process because, in the case of magnitude, the possibility to represent the hidden variable and the observations in the same space allows the learner to reason about them using psychological and neural mechanisms specific to that space (for example, integrating stimulus positions/numerosities to calculate the average position/numerosity).

Another difference between magnitude learning and probability learning is the degree of informativeness of an observation for learning the hidden variable. The observations are typically much less informative for learning probabilities than for learning magnitudes. This is covered in Chapter 2, and in that chapter I show that this has consequences for the learning process.

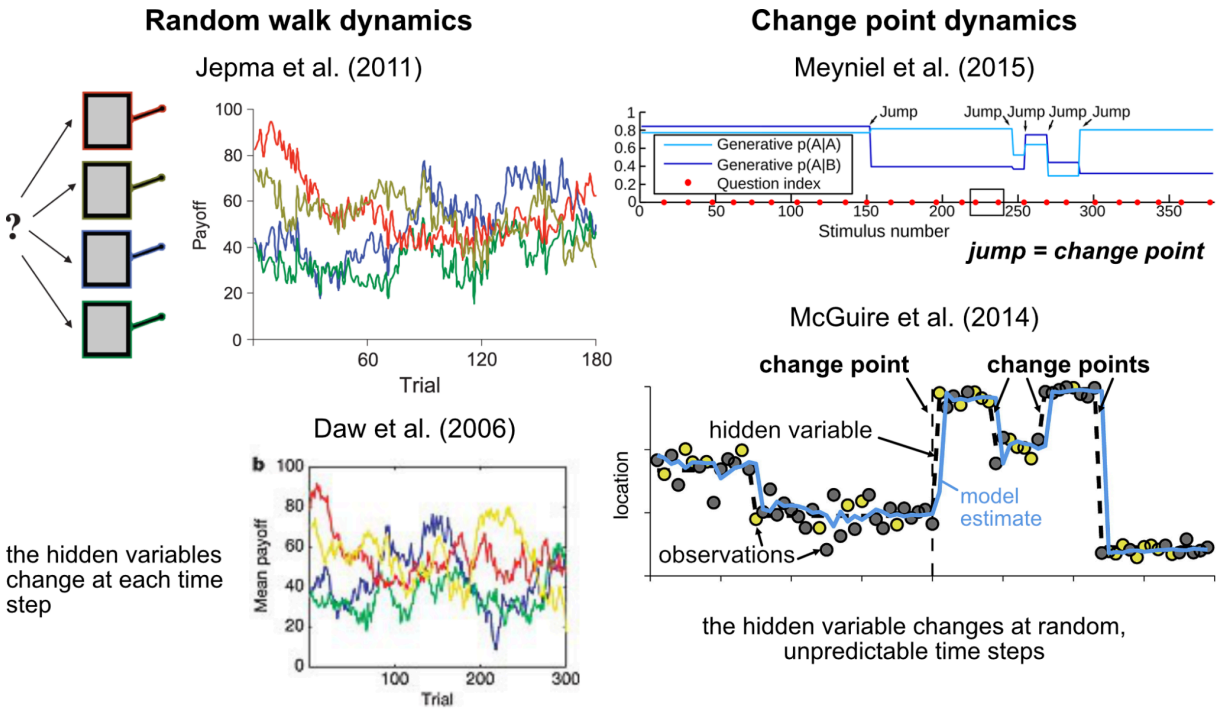


Figure 3. Types of dynamics. Panels adapted from (Daw et al., 2006; Jepma & Nieuwenhuis, 2011; McGuire et al., 2014; Meyniel et al., 2015).

Two types of dynamics can be studied: ***random walk*** dynamics, where changes occur predictably at each time step, or ***change point*** dynamics, where changes occur at unpredictable time steps called *change points* (**Figure 3**). In the case of random walk dynamics, the hidden variable changes at each time step by an independent random amount. The variance of those amounts, which dictates the average amplitude of the change in hidden variable from one time step to the next, is called ***volatility***. In the case of change point dynamics, at each time step, a change point can occur randomly with a certain probability. When a change point occurs, the hidden variable changes abruptly, otherwise, it remains the same. The probability that a change point occurs is referred to as ***probability of change point***, or ***volatility***—*please note* that the volatility in the case of change point dynamics is ***not*** the same as the volatility in the case of random walk dynamics—, or *hazard rate* (this term has been borrowed from the engineering literature used because historically change points modelled failures in plants or engineering components). The random walk and change point dynamics will be described formally and more thoroughly in section 3.1.

A key difference between these two types of dynamics is that in the random walk case, the times at which changes occur are *known* (they occur on every trial), whereas in the change point case, the times at which changes occur are *unknown* and *unpredictable* (these are the change points). In the latter, the learner must try to infer when a change has occurred to perform the task effectively. This *change point inference* process is only called upon during learning for the second type of dynamics, and poses unique challenges as we will see later.

Two types of responses can be required from the subject: based on what they have learned, they must either make a ***choice*** between several options, or provide an ***estimate*** on a continuous scale of a hidden variable (**Figure 4**). The choice case can further be splitted into two subcases depending on whether it is a perceptual choice or a reward-based choice. In the perceptual case, as in (Glaze et al., 2015) shown in **Figure 4** and in (Murphy et al., 2021; Ossmy et al., 2013), the subject chooses, having observed a certain number of perceptual evidence samples (the observations $x_{1:t}$, star positions in Glaze et al.), which distribution is the one currently generating the samples, among a set of possible distributions (each distribution corresponds to one possible value of h_t, which is categorical in this case; the two possible distributions are represented by the two triangles in Glaze et al.). In the case of reward-based choices, typically in multi-armed bandit tasks as in (Daw et al., 2006; Findling et al., 2019; Jepma & Nieuwenhuis, 2011; Pulcu & Browning, 2017) shown in **Figure 2**, **3** and **4**, at each time step, the subject makes a choice and receives a reward (the observation x_t, e.g. in Findling et al. the number of points of 52 displayed on Figure 4 and 2) which is sampled according to the reward distribution of the option they have chosen. From the observed rewards, the subject must learn the parameters of the reward distributions of the different options (the hidden variable h_t, e.g. in Findling et al. these are the mean rewards associated with each of the two possible choices) in order to maximise their cumulative total rewards. In both the perceptual and the reward-based cases, the subject's choice is informed by what they have learned from past observations, which allows researchers to make inferences about their underlying learning process, but this inference is rather indirect. The estimate case is more straightforward: based on the observations they have been presented with ($x_{1:t}$), the subject provides a continuous estimate (v_t) of the hidden variable (h_t), typically using a slider as shown on **Figure 4**.

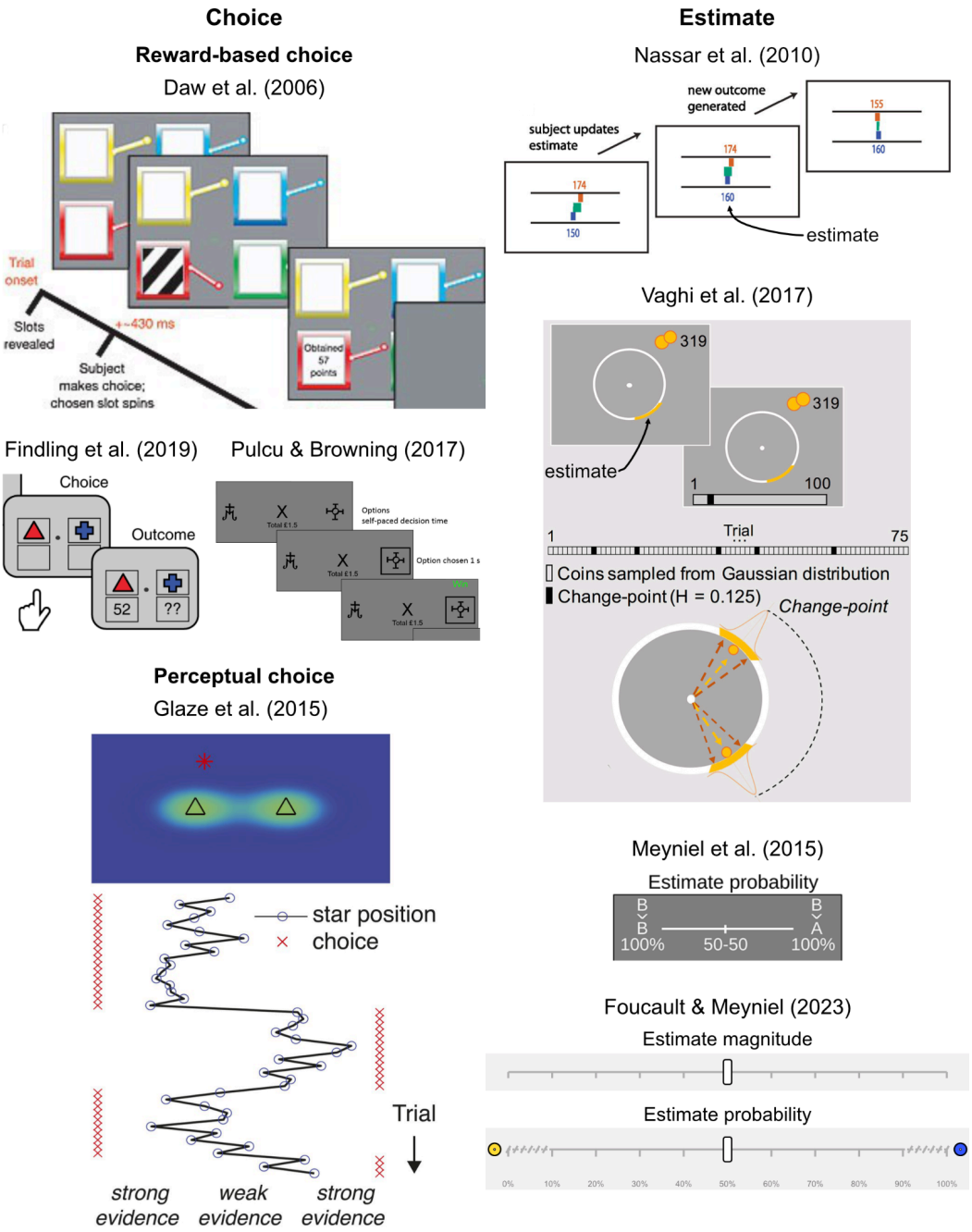


Figure 4. Types of responses. Panels adapted from (Daw et al., 2006; Findling et al., 2019; Foucault & Meyniel, 2023; Glaze et al., 2015; Meyniel et al., 2015; Nassar et al., 2010; Vaghi et al., 2017).

1.5. Adaptive learning at different levels

Here, we will look at the challenges of adaptive learning in stochastic and dynamic environments, and distinguish two levels of adaptive learning with specific challenges whose traces in human behaviour will be explored in section 2.

In a stochastic and dynamic environment, effective learning requires adapting the learning rate because there is a permanent tension between stochasticity and the dynamics, which encourage the learner to decrease and increase their learning rate, respectively (Soltani & Izquierdo, 2019). Let's illustrate in turn the influence of stochasticity and the influence of the dynamics.

Stochasticity introduces variations in observation values that occur even when the underlying hidden variable is held fixed. This means that a single observation is not sufficient to obtain an accurate estimate of the hidden variable; multiple past observations are needed, which corresponds to using a learning rate lower than 1. The greater the stochastic variations, the more advantageous it is for the learner to reduce their learning rate as this will average observation values across a larger window and thereby increase the precision of the estimate, assuming that the hidden variable has not changed within that window. However when a change in hidden variable occurs, using a low learning rate is disadvantageous. Indeed, if the learner knew for sure that a change had just occurred and that their previous knowledge was irrelevant for knowing the new hidden variable value, they would need to discard their old estimate and base their new estimate only on the upcoming observation. This corresponds to using a learning rate equal to 1. In practice, changes are hidden, so the learner is never certain whether a change has occurred (when the dynamics have unpredictable change points), or how much change there has been. The general principle is that the more likely a change has occurred (or the more likely changes are to be large), the higher the learning rate should be.

The tension between the ability to deal with stochasticity versus the dynamics is sometimes referred to as the stability-flexibility trade-off (Nassar & Troiani, 2021). The value of the learning rate corresponds to a certain value of the trade-off: the higher the learning rate, the more flexibility is favoured over stability.

In this thesis, ***adaptive learning*** generally refers to a learning process that features an adaptation of the learning rate ("generally" in the sense that the ability to adapt the learning rate is relevant to all chapters of the thesis; in chapter 3 specifically additional adaptive capabilities are also explored).

Two levels of adaptive learning can be distinguished. The **first level** concerns the *average learning rate* and its adaptation based on the average level of stochasticity or volatility in the environment. Average learning rate is contrasted with instantaneous values of the learning rate, which are the focus of the second level. The **second level** involves *dynamically adapting the learning rate* for each observation in order to better respond to a change and disregard stochastic fluctuations. Let's examine in more detail what these two levels entail—I informally describe below the adaptation principles that are called for in each level, and they will be characterised more formally (mathematically and computationally) in section 3.

Level 1 of adaptive learning. There are two kinds of adaptation at level 1: adaptation to stochasticity and adaptation to volatility.

> *Adaptation to stochasticity:* When the level of stochasticity in the environment is higher, the learning rate should be lower on average to compensate for the loss of precision induced by greater stochasticity. Precision is gained by taking into account a larger number of past observations for the estimation.
>
> *Adaptation to volatility:* When the volatility of the environment is higher, the learning rate should be higher on average. Volatility quantifies the average rate of change; a higher volatility means that changes in the hidden variable occur on average more frequently (when the dynamics are of the change point type), and/or have higher average amplitude. A higher learning rate when volatility is higher is adaptive allows for a quicker convergence of the estimate to the new value of the hidden quantity each time a change occurs.

What is normatively relevant for calibrating the average learning rate in level 1 is the relative level of volatility compared to stochasticity. However, it is useful from an empirical point of view to consider volatility and stochasticity separately as they are two independent quantities.

Level 2 of adaptive learning. Level 2 corresponds to dynamic adjustments of the learning rate, i.e. made on an observation-by-observation basis, and with different adjustments depending on what is observed (not simply on the observation index, in which case the adjustment process is not truly dynamic in the sense that learning

rates could be determined in advance). These dynamic adjustments are typically called for in the case of change point dynamics: the learning rate should ideally be increased when a change point occurs, i.e. when the hidden variable has changed, and decreased after the change point when only stochastic fluctuations occur. As change points are hidden, these adjustments should be based on the probabilities inferred from the observations received that a change point has occurred in the recent past (i.e., the probability that a change point has occurred at time *t*, at time *t-1*, ...). In that case, the dynamic adjustments are the result of a change point inference process.

Note that the level 2 dynamic adjustments of the learning rate in response to individual change points imply a level 1 adaptation of the average learning rate to the average rate of change points: when change points are more frequent, transient increases in the learning rate occur more frequently, which is reflected overall by an increase in the average value of the learning rate.

In the literature, and in this thesis, level 2 of adaptive learning is primarily studied in the case of change point dynamics. However, it could in principle also be studied in the case of random walk dynamics, assuming that volatility and/or stochasticity can vary dynamically in the environment. In that case, the learner should infer, based on the received observations, the current relative level of volatility compared to stochasticity, and dynamically adjust their learning rate accordingly. This case is connected to the change point case: One can think of the change point case as an extreme case of dynamically varying volatility/stochasticity ratio in a random walk, where the volatility/stochasticity ratio has a base level of 0 and occasional instantaneous spikes (which creates change points).

2. Behavioural findings on adaptive learning

We have seen that there are several types of challenges and several levels of adaptive learning. Here, we will examine what has been discovered about how humans meet these challenges behaviorally. We will see that numerous findings

have been made about level 1 but few about level 2 of adaptive learning, and that the latter remains unexplored in probability learning.

The behavioural findings in this section are organised into two parts, corresponding to the two levels of adaptive learning. The first part focuses on the average measure of learning rate across a block of trials, and encompasses the majority of studies. The second part, which has been examined in a narrower range of learning contexts, involves measuring the learning rate on an observation-by-observation basis. The asymmetry between the two parts is attributable to the fact that in many experimental paradigms measuring the learning rate on each observation is not feasible. In order to calculate the learning rate using equation **[1]**, a continuous measure of v_t (the subject's estimate) is needed, which excludes all studies where the subject's responses are choices, or are not provided on each observation.

2.1. Behavioural results related to the average learning rate (level 1)

The results regarding the adaptation of the average learning rate can be divided according to the two main effects of environmental volatility and stochasticity. In the following paragraphs, I discuss the results related to these two effects separately, detailing the different learning contexts in which they have been shown.

Effect of volatility

A highly influential study by Behrens et al. demonstrated an effect of volatility on average human learning rate in a probability learning, choice-based, reinforcement learning context (Behrens et al., 2007). The task is a probabilistic reversal learning task, a type of two-armed bandit task like that of (Pulcu & Browning, 2017) shown on **Figure 2** (but with only one type of outcome rather than two). In each trial, subjects choose one of two bandit arms and either receive or do not receive a reward based on the reward probability p_t, which is the hidden quantity that the subject needs to learn, p_t=h_t. One arm has a reward probability of p_t, while the other has a reward probability of 1-p_t. The volatility in this task is manipulated using two conditions, each corresponding to a different block of trials (120 or 170 trials) that the subject performs consecutively. In the stable condition, the reward probability remains the same throughout the block of trials (p_t=75%, volatility=0). In the volatile condition, the

reward probability undergoes reversals (alternating between p_t=80% and p_t=20%) regularly every 30 or 40 trials (volatility=1/30 or 1/40). The key question is whether the subjects' average learning rate in the volatile block differs from that in the stable block. The method used in this type of task to measure the average learning rate is to fit a model—see **Box 2** for an explanation.

The results obtained by Behrens et al. show that the average learning rate of the subjects in a volatile block is significantly higher than that in a stable block (**Figure 5**). These results have been replicated in subsequent studies (although the size or significance of the effect are not always replicated) and generalised to cases where the valence of reinforcements is negative (e.g. electric shocks or money losses), and cases where two types of outcomes occur on each trial according to two distinct hidden probabilities (Browning et al., 2015; Cook et al., 2019; Pulcu & Browning, 2017).

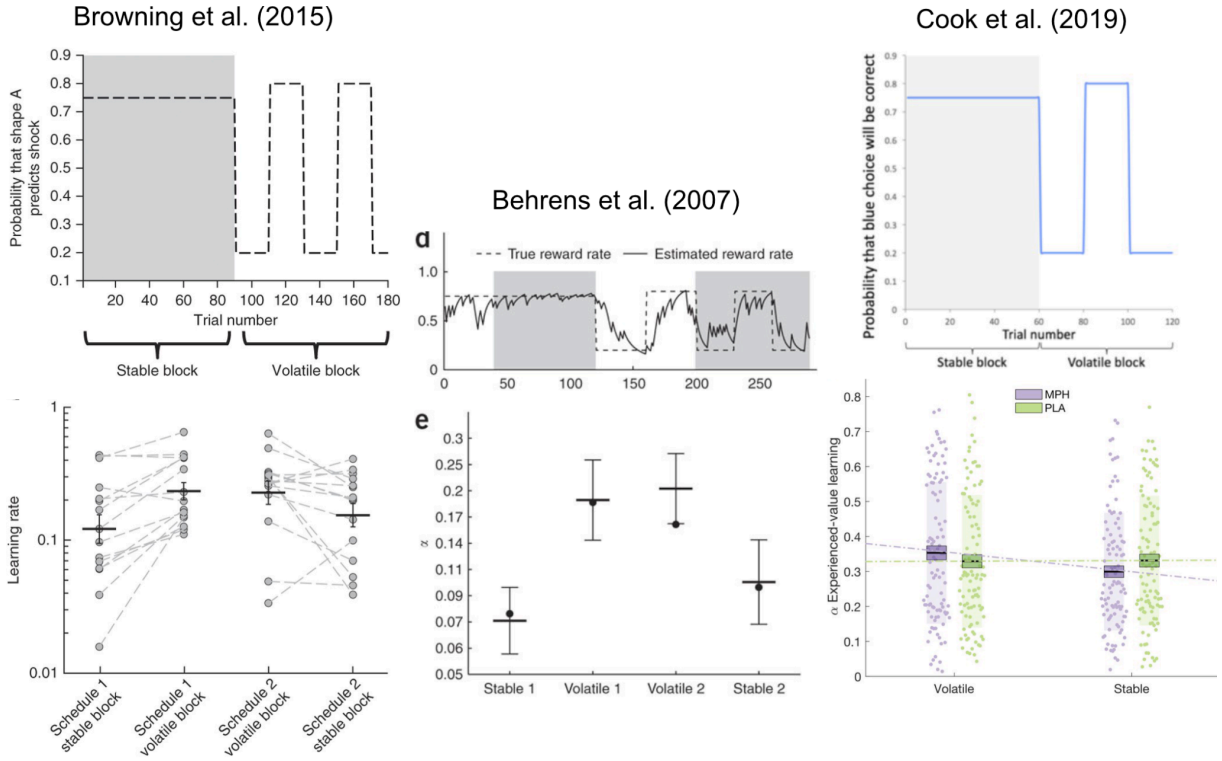


Figure 5. Average learning rate of subjects in a volatile block compared to a stable block in reinforcement learning tasks. Panels adapted from (Behrens et al., 2007; Cook et al., 2019; Pulcu & Browning, 2017).

Box 2. Measurement of average learning rate through model fitting.

In choice tasks, the typical method used to measure the average learning rate is to use a model that comprises a learning rate parameter and fit it to the subject's choices. The fitted parameter value is used as a measure of the average learning rate. When comparing the average learning rate in two conditions (e.g., stable vs volatile blocks), the learning rate is fitted twice separately on the blocks of trials corresponding to the given condition.

In most studies, the learning rate parameter is involved in a delta-rule used as a learning component of the model (Bush & Mosteller, 1951; Rescorla & Wagner, 1972; Sutton & Barto, 2018; Widrow & Hoff, 1960). This delta-rule assumes a fixed learning rate and is described by the update equation **[2]** below:

$v_t = v_{t-1} + \alpha\,(x_t - v_{t-1})$ **[2]**

The delta-rule is integrated within a larger model that includes a choice component, such as Q-learning or SARSA or another reinforcement learning model (Sutton & Barto, 2018), in order to model the subject's choices. The fitting process involves comparing the subject's choices to those that the model would have made under a given learning rate parameter value, and adjusting the learning rate parameter so that the subject's choices and the model's choices are as close as possible. Technically, the learning rate is fitted jointly with the other parameters of the model, and the process of adjusting the parameters corresponds to an optimization process that aims to minimise a cost function. In a probabilistic modelling framework, following the maximum likelihood principle, the canonical cost function is the negative log-likelihood of the choices made by subjects according to the model (Goodfellow et al., 2016; Wilson & Collins, 2019).

The effect of volatility has also been studied in non reinforcement-based learning tasks, which are perceptual evidence accumulation tasks, like the one by (Glaze et al., 2015) shown in **Figure 4**. These studies have been conducted notably in the laboratories of Joshua Gold and Tobias Donner (Glaze et al., 2015; Murphy et al.,

2021; Ossmy et al., 2013). In this type of task, the hidden variable that the subject needs to learn is a binary state variable indicating which of two distributions is the current generative distribution of the observed evidence samples. Each sample constitutes an observation. In Glaze et al's study, for example, each observation is the position of a star on the screen, and the two possible distributions generating the star positions are two Gaussian distributions centred at opposite sides of the screen.

As a behavioural report, the subject must choose which of the two distributions they believe to be the current generative distribution of the samples. At unpredictable change points, the true hidden state switches, with a certain hazard rate. Different hazard rate values are used in different blocks of trials. For example, in Glaze et al.'s study, subjects performed blocks of 1000 trials, each block having a hazard rate chosen from a set of 7 possible values (0.05, 0.1, 0.3, 0.5., 0.7, 0.9, 0.95).

As in the previously-mentioned reinforcement learning tasks, the effect of volatility on learning was studied by fitting a model separately in the different blocks corresponding to different volatility conditions. The models used in perceptual evidence accumulation studies were different from those used in reinforcement learning studies and did not have a learning rate parameter, but another parameter that plays an analogous role to the average learning rate: the subjective hazard rate. The results of these studies show that the subjective hazard rate of the subjects is consistently adapted based on the objective hazard rate of the considered block (Glaze et al., 2015).

Effect of stochasticity

Compared to volatility, the effects of stochasticity have been shown in a smaller number of studies, with less known generality. These studies focus on a magnitude learning context (rather than a probability learning context).

One study where the effect of stochasticity on the learning rate is most clearly presented is the study by Lee et al. (Lee et al., 2020). In this task, as in other magnitude learning tasks, the subject needs to estimate the mean of the observation generation distribution (a Gaussian distribution), which is the hidden quantity. The observation values are continuous (in Lee et al.'s task they correspond to stimulus

positions). After each observation, the subject reports their estimate of the hidden quantity (formulated in Lee et al.'s task as an estimate of the predicted position of the next stimulus). The hidden quantity is also dynamic; in Lee et al.'s study these dynamics followed a Gaussian random walk process (described in section 3.1). Stochasticity is quantified by the standard deviation (SD) of the observation generation distribution (the larger the SD, the higher the stochasticity).

To study the effect of stochasticity, Lee et al. used two conditions: one with an SD of 10 and the other with an SD of 25, on a domain of position values ranging from 0 to 300 (0 and 300 correspond to the leftmost and rightmost possible stimulus position, respectively). Lee et al. measured the subjects' average learning rate in each condition from their estimate reports. The results show an effect of stochasticity in the expected direction: the subjects' average learning rate is significantly higher in the condition with lower stochasticity than in the condition with higher stochasticity (**Figure 6**). The difference in subjects' learning rates was however smaller compared to that predicted by the optimal model: the average learning rates were 0.59 and 0.69 in the two conditions for the subjects, vs. 0.35 and 0.65 for the optimal model.

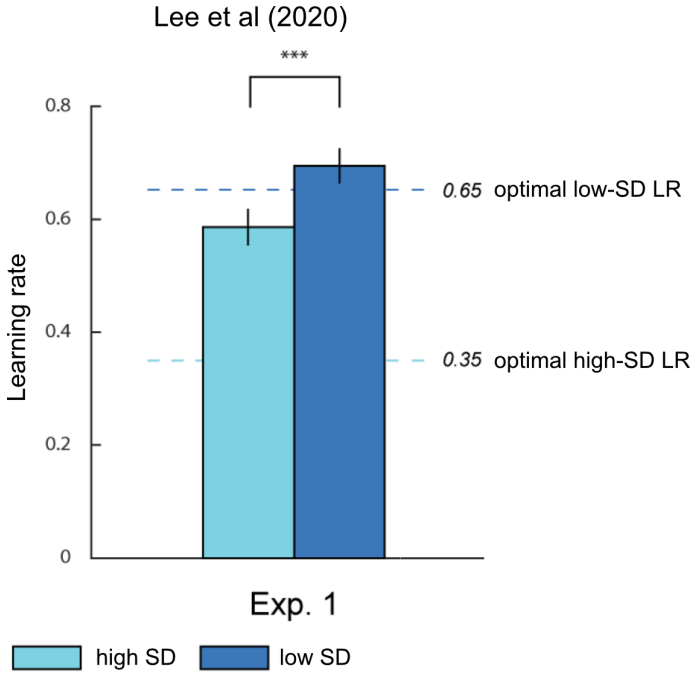


Figure 6. Average learning rate of subjects in a condition with high stochasticity compared to a condition with low stochasticity. Figure adapted from (Lee et al., 2020).

Stochasticity has also been manipulated in magnitude learning studies with change point dynamics, but the effects of the manipulation of the SD on subjects' average learning rate have not been reported as clearly as in Lee et al.'s study. (Nassar et al., 2010) used four SD conditions and reported a plot of the subjects' learning rate as a function of the absolute value of the prediction error for each condition. These plots visually show lower learning rate values overall in conditions with higher SD, but statistical tests have not been reported. (McGuire et al., 2014) used two SD conditions and reported in the supplementary materials an analysis showing that the modulation of subjects' learning rate by the absolute prediction error was significantly modulated by the level of SD. This effect was mainly manifested for intermediate values of prediction error, because normatively an intermediate value of prediction error indicates a high change-point probability in a low SD condition but a low change-point probability in a high SD condition, and the change-point probability in turn modulates the learning rate. Overall, it seems that most of the effect of stochasticity on the subjects' learning rate in change-point magnitude learning tasks is mediated by the change-point probability. This is consistent with the Chapter 2 results showing that most of the variations in the learning rate is explained by the change-point probability in magnitude learning tasks.

2.2. Behavioural results related to the dynamic adjustments of the learning rate (level 2)

Previous studies on the dynamic adjustments of the learning rate at each observation have been restricted to one class of learning tasks: magnitude learning tasks, with change point dynamics, and estimate reports. These studies have been conducted in particular by Matthew Nassar and colleagues. They have shown that, in magnitude learning, humans dynamically adapt their learning rate by transiently increasing it following a change point (see **Figure 7**) (McGuire et al., 2014; Nassar et al., 2012; Vaghi et al., 2017). The increase is immediate (occurring at the first observation following the change point) and relatively short-lived (the increase is mainly concentrated on the first outcome, almost back to baseline by the third outcome). In Chapter 2, I will demonstrate that this "one-shot" profile of dynamic adjustment is characteristic of magnitude learning tasks.

Nassar et al.'s results have been replicated and generalised across different types of magnitudes, including symbolic numbers (Nassar et al., 2012), positions (McGuire et al., 2014), and angular directions (Vaghi et al., 2017). Note that, as mentioned earlier in section 1.5, these level 2 findings also imply that the average learning rate should increase with volatility, further extending the volatility findings covered in section 2.1.

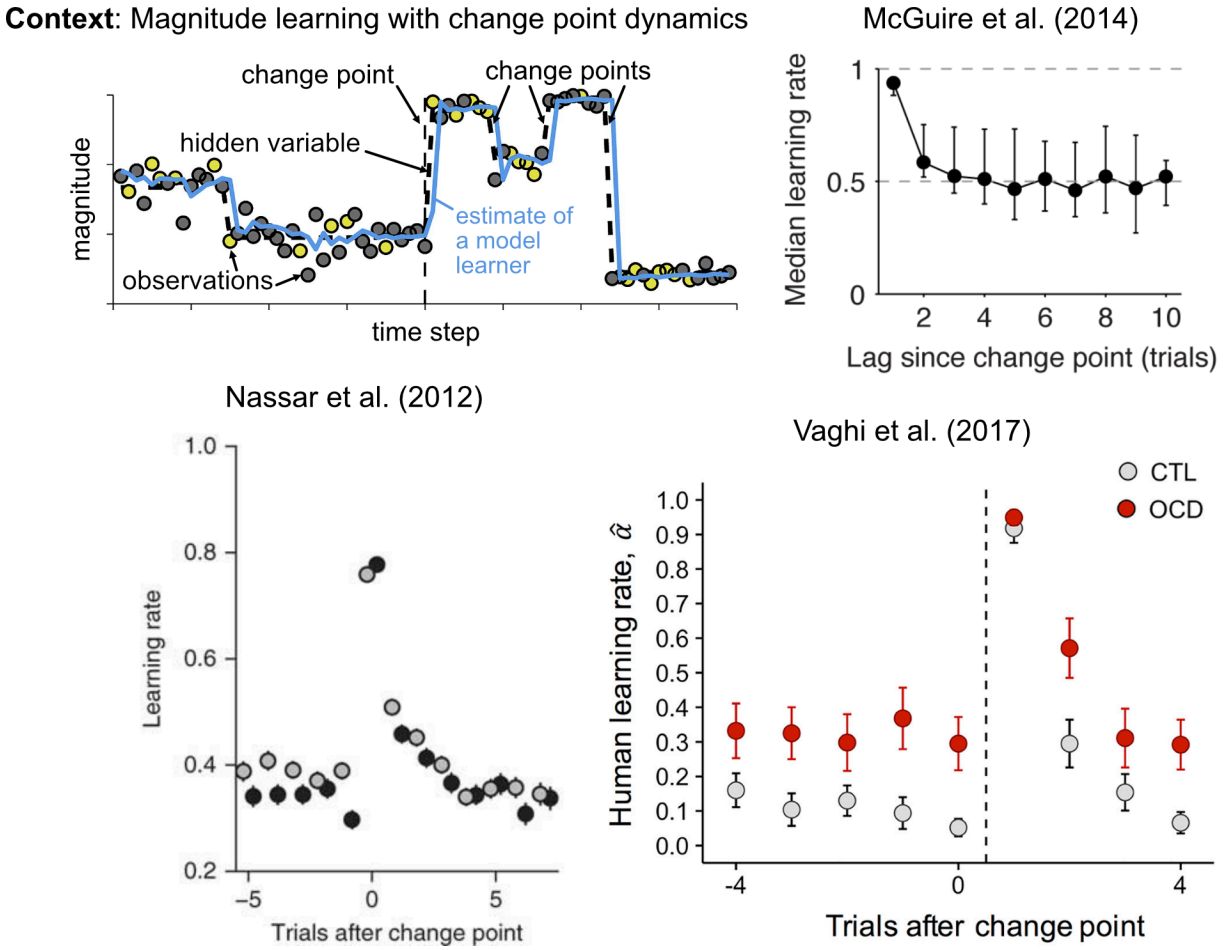


Figure 7. Learning rate of subjects following a change point during magnitude learning. Panels adapted from (McGuire et al., 2014; Nassar et al., 2012; Vaghi et al., 2017).

The other analyses reported in these magnitude learning studies show that human learning rates increase with the absolute value of prediction error (consistent with the fact that larger prediction errors indicate a higher change-point probability) and are significantly modulated by change-point probability and relative uncertainty, as estimated by a normative model of the task (more details about normative models will be provided in section 3).

To my knowledge, the study by (Nassar et al., 2010) was the first to collect behavioural measures of the apparent learning rate using equation **[1]**. Unlike the method described in **Box 2**, this measure does not depend on any modelling assumptions, and allows for studying how the learning rate is dynamically adjusted in response to task events and as a function of dynamic factors. These are advantages that I fully exploit in Chapters 2 and 3 by similarly collecting observation-by-observation estimate reports and using the measure of the apparent learning rate.

2.3. Open question: Dynamic adjustments in probability learning, and comparison with magnitude learning

A gap in scientific knowledge concerns the dynamic adjustments of human learning rates in probability learning. As covered in section 2.1, there is ample empirical evidence that the human learning rate is on average adapted depending on whether the environment is stable or volatile (level 1 of adaptive learning). But is the human learning rate dynamically adapted on the basis of each observation in response to a single change in the environment (level 2 of adaptive learning)?

Examining these dynamic adjustments in probability learning is necessary because, firstly, there is a fundamental difference in the informativeness of the observations in these two types of learning, as detailed in chapter 2, and in fact, I show in chapter 2 that this leads to differences in the learning process in these two contexts. Secondly, the results presented in section 2.1 regarding probability learning can be explained in several ways that do not involve level-2 dynamic adaptation. Subjects could, for example, exhibit two different average learning rates in low and high volatility conditions by adopting two different behavioural strategies and switching between the two depending on the condition they are in, without each strategy involving any dynamic adaptation of the learning rate. Two such strategies that often explain human choices well are the win-stay, lose-shift strategy (well-suited when volatility is high), and a strategy with a strong choice repetition bias (well-suited when volatility is low) (Cook et al., 2019; Palminteri, Wyart, et al., 2017; Wilson & Collins, 2019). Another possibility would be that the subjects' learning rate gradually increases or

decreases over time depending on the average number of change points over a long period of time, without exhibiting any adaptation in response to a single change point.

The need to examine the dynamic adjustments of the learning rate in probability learning and the lack of suitable data for this examination were one motivation for the study presented in chapter 2. Another motivation was the need to bring together and compare magnitude learning and probability learning, which, despite being two very common types of learning, have been studied by largely separate branches of the literature (cf. **Table 1**). How do they relate to each other? What do these two types of learning have in common, and what are their differences?

3. Computational bases of adaptive learning

In this section, I discuss the computational process involved in adaptive learning from a normative perspective. I formally describe the problem posed in learning and derive the optimal solution to this problem. I then identify the properties that govern adaptive learning in this solution. These normative properties will define desiderata and evaluation criteria for computational models of learning. Next, I will summarise the computational models proposed in the literature to solve the learning problem, and will then highlight key limitations of previous approaches and outstanding questions that I will address in chapter 2 using a neural network-based approach to investigate the feasibility and realisation of adaptive learning in the brain.

The normative approach which I adopted here is useful for several reasons. First, the optimal solution allows me to describe how the learning rate should be adapted and to identify computational factors that modulate the adaptations in the optimal solution, and that should be taken into account in any other adaptive solutions. This allows me to formulate general principles governing adaptive learning. Additionally, the optimal solution serves as a benchmark for any other learning model. The quality of the solution provided by a model, and of its adaptive capacities, can be quantified in comparison to the optimal solution. Finally, the normative approach allows me to make empirical predictions that explain both "how" learning is performed and "why" it is performed in this way (because it follows from an optimality principle), and that

have the potential to generalise beyond a specific experimental situation (because the ingredients from which these predictions are derived, namely the optimality principle, probability theory, and the generic learning problem I describe, have broad applicability). Normative predictions thus have strong descriptive value, strong interpretive value, and strong potential for generalisation (Courville et al., 2006; Dayan et al., 2000).

3.1. Modelling the learning problem

In all the contexts studied here (**Table 1**), the task requires solving a learning problem. This problem can be asked to the subject directly (in estimation tasks, the subject must report the learned value) or indirectly (in choice tasks, choices are made based on the learned values).

Formally, the learning problem corresponds to a sequential probabilistic inference problem. Probabilistic inference consists of estimating (as accurately as possible) the hidden variable of the environment from stochastic observations. The inference is probabilistic due to the stochasticity (it is impossible to determine with certainty the exact value of the hidden variable from the observations). In the context of learning, the inference is sequential because the observations are received sequentially, and because we are interested in the updating process.

Specifying the problem requires specifying the hidden variable to be estimated, and how they are related to the observations in the considered environment. This specification is mathematically modelled by a stochastic process called the ***generative process*** of the environment. The process is referred to as *generative* because it generates the observation sequence.

One way to describe the generative process involved in the learning problem in a generic way is through a probabilistic graphical model (also called Bayesian network, and here more specifically a dynamic Bayesian network, because it relates variables over adjacent time steps), which is the one shown in **Figure 1**. The graphical model represents the conditional dependence structure between the random variables of the process. Each edge represents a direct conditional dependency. Any variable in the graph is conditionally independent of all its non-descendants given the value of

all its parents. For example, h_{t+1} is conditionally independent of x_t (a non-descendent) given h_t (its parent). This conditional dependence structure applies to all learning environments, and is key to solving the learning problem, as we will see in the next section. (Note that this structure is not as restrictive as it might seem in fact, because a generative process that is described by a graph that does not have this structure can often be turned into an equivalent one that does have this structure, by a change of variable in which the new hidden variable contains the conditional dependencies of the original graph that are not captured in the structure from Figure 1.)

To fully specify the generative process, one only needs to specify for each variable *z* the probability distribution for *z* conditional upon *z*'s parents in the graph. Thus, in the case of learning, to define the generative process of a particular studied environment, only three probability distributions need to be defined.

1) The initial distribution of the hidden variable: $p(h_0)$.

2) The dynamics distribution: How the next value of the hidden variable (h_{t+1}) is generated based on the previous value (h_t). This is the distribution: $p(h_{t+1} \mid h_t)$.

3) The observation generation distribution: How an observation (x_t) is generated given the values of the hidden variable (h_t). This is the distribution: $p(x_t \mid h_t)$.

Definition of the generative process for specific learning contexts

Three common learning contexts are magnitude learning with random walk dynamics, magnitude learning with change point dynamics, and probability learning with change point dynamics (**Table 1**). These correspond to distinct classes of environments. Detailing the specificities of each of these classes is useful because their specific properties have important computational consequences. The classes of environments corresponding to each learning context are defined by the generative processes below.

The case of change point dynamics is subdivided into two cases that commonly occur in experiments: either **(a)** the hidden variable *varies uniformly* over an interval

(e.g. (Gallistel et al., 2014; Nassar et al., 2010)), or **(b)** the hidden variable makes *reversals* between two values (reversal tasks) (e.g. (Behrens et al., 2007; Glaze et al., 2015)).

Magnitude learning with random walk dynamics.

1) Initial distribution of the hidden variable. This is typically a uniform distribution over the domain of the hidden variable (for magnitude: a continuous range of magnitudes, whose bounds depend on the application domain), or a Gaussian distribution around the central value of the domain.

2) Dynamics distribution. The hidden quantity evolves according to a random walk process. At each time step, a random step (i.e. amount of change) is drawn from a distribution and injected into the hidden quantity. Importantly, these steps are independent. This random walk process is characterised by the step distribution, which is typically Gaussian, and the standard deviation of the distribution, σ_h, which defines a volatility level (typically experimental-controlled). It may also involve a decay (parameterised by a decay rate λ and a convergence value h_∞) but most often in experimental studies the decay is absent (λ=1) or very small. For a Gaussian random walk, the dynamics are thus described as:

$h_{t+1} \sim N(\lambda\, h_t,\, \sigma_h^2)$, or equivalently: $h_{t+1} = \lambda\, h_t + \varepsilon_{t+1}$ where $\varepsilon_{t+1} \sim N(0,\, \sigma_h^2)$

3) Observation generation distribution. In the case of magnitude learning, the hidden variable and the observations belong to the same domain, which is a domain of magnitudes. The hidden variable defines the mean of the observation generation distribution, which is typically Gaussian (or a circular normal distribution, for circular magnitudes), and characterised by a standard deviation, σ_x, which defines a stochasticity level (typically experimental-controlled). Thus, the observation generation distribution is $N(h_t, \sigma_x^2)$, i.e.:

$x_t \sim N(h_t,\, \sigma_x^2)$

Magnitude learning with change point dynamics.

1) Initial distribution of the hidden variable. Case **(a):** A uniform distribution is taken over the domain of the hidden variable. Case **(b):** An equal probability for the two values between which the hidden variable reverses.

2) Dynamics distribution. It is characterised by a probability p_c that a change point occurs (which is the volatility level, typically experimenter-controlled), and by a distribution for resampling the hidden variable when a change point occurs. The dynamics from h_t to h_{t+1} are described as:

> "*with probability p_c, h_{t+1} is sampled from the resampling distribution, otherwise, $h_{t+1}=h_t$*".

The resampling distribution is different in case **(a)** and case **(b)**. In case **(a)**, the resampling distribution is equal to the initial distribution, meaning that after the change point, the value of the hidden variable is uniformly resampled over its domain (possibly with some constraint to ensure that the magnitude of the change is not too small). In case **(b)**, after the change point, the value of the hidden variable reverses.

3) Observation generation distribution. Same as for magnitude learning with random walk dynamics.

Probability learning with change point dynamics.

1) Initial distribution of the hidden variable. Same as for magnitude learning except that the domain of the hidden variable is always bounded between 0 and 1, since the hidden variable is a probability.

2) Dynamics distribution. Same as for magnitude learning with change point dynamics.

3) Observation generation distribution. In the case of probability learning, the observations are categorical and the hidden variable defines the probability of each category. In the simplest, most common case, the observations are binary, and each observation is sampled from a Bernoulli

distribution whose parameter is equal to the hidden variable. Thus, the observation generation distribution is $Bern(h_t)$, i.e.:

$$x_t \sim Bern(h_t)$$

A more complex case of probability learning, which I study in Chapter 3, is the case where observations are generated according not to a single hidden probability *p(1)* but to two hidden probabilities *p(1|0)* and *p(1|1)* (Foucault & Meyniel, 2021; Heilbron & Meyniel, 2019; Meyniel et al., 2015). In this case, the observation is still generated according to a Bernoulli distribution, but the Bernoulli parameter is either one or the other of the two hidden probabilities, depending on the previous observation. This also introduces an additional level of complexity in the dynamics because the change points of the two hidden probabilities can occur either independently or at the same time, which in the latter case introduces coupling between the two hidden probabilities and an additional level of hierarchy in the inference process (see Chapter 3).

3.2. Optimal solution

The learning problem described above, once specified using probability distributions, has an optimal solution that can be derived from probability theory. This solution uses the principle of Bayesian inference (Ma et al., 2023). More importantly for learning, the optimal solution can be computed through a sequential update process, where the next belief is calculated based on the previous belief and the observation received at time t (rather than being calculated from scratch each time by considering all the observations received from the start, as a direct application of Bayes' rule would dictate). This sequential solution is called *Bayesian Filtering* or *Recursive Bayesian Estimation* (Chen, 2003; Särkkä & Svensson, 2023). I provide below the essential elements of this optimal sequential solution.

The problem requires estimating the hidden variable, h_t, given the observations received so far, $x_{1:t}$. The solution prescribed by probability theory is to calculate the posterior probability distribution over the hidden variable given the observations, $p(h_t \mid x_{1:t})$, which I will simply refer to as *the posterior*, and from the posterior, to calculate a point estimate of the hidden variable. The optimal point estimate depends on the

exact task at hand. Without loss of generality, I will consider in the following that the estimate is taken as the mean of the posterior, which is optimal, for example, to minimise the mean squared error between the estimate and the true value of the hidden probability. It would be easy to swap the mean for another statistic such as the median (which minimises the mean absolute error) or the mode (maximum a posteriori estimate). The core of the problem is computing the posterior.

According to Bayes' rule, the posterior is equal to:

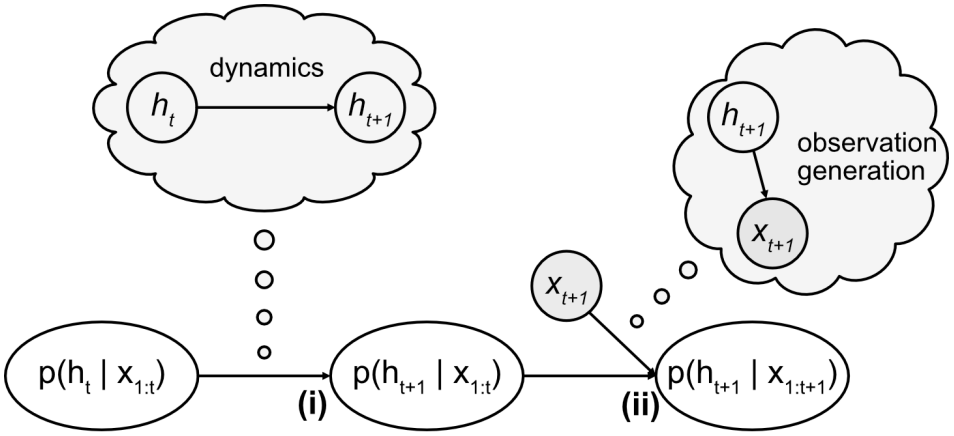
$$p(h_t \mid x_{1:t}) = p(x_{1:t} \mid h_t)\, p(h_t) \,/\, p(x_{1:t}) \quad [3]$$

Unfortunately, this solution is not very useful, for two reasons. The first reason is that the terms on the right-hand side of equation [3] cannot be directly calculated from the generative process distributions. To compute them, one would first need inject all past values of the hidden variable into equation [3] in order to compute the *joint* distribution $p(h_{1:t} \mid x_{1:t})$ rather than $p(h_t \mid x_{1:t})$. This would then allow one to calculate the first term of the right-side as a product of $p(x_i \mid h_i)$ for all i, each of which can be calculated using the observation generation distribution, and to calculate the second term $p(h_{1:t})$ as a product involving the initial distribution $p(h_0)$ and the dynamics distribution $p(h_i \mid h_{i-1})$ for all i. Finally, having computed the joint distribution, one would then need to marginalise all previous values of the hidden variable ($h_{1:t-1}$) in order to obtain the desired distribution $p(h_t \mid x_{1:t})$.

The second, more fundamental reason why this solution is not so useful as a model of the learning process involved in animals and brains, is that it has the major disadvantage that each time a new observation is received, the posterior must be recalculated from scratch, taking into account all the observations received from the start, $x_{1:t}$. This is not only computationally costly, but also not very relevant from a psychological and neuroscience point of view because the brain cannot store in memory all the observations received over an arbitrarily long period of time, and because in these disciplines the learning process is conceived of as an update process.

For these reasons, a sequential method for calculating the posterior would be more useful. One such sequential method is called ***Bayesian filtering***, also known as

Recursive Bayesian estimation (Chen, 2003; Särkkä & Svensson, 2023). It is applicable whenever the generative process has the conditional independence properties described on **Figure 1**. It computes the successive posteriors following each observation without retrieving all previous observations, using instead the posterior previously computed. It thus presents the calculation of the posterior as a sequential update process, which is more compatible with learning in the sense of psychology and neuroscience. It is described by the algorithm below, graphically illustrated in **Figure 8**.



(i) $p(h_{t+1} \mid x_{1:t}) = \int p(h_{t+1} \mid h_t)\, p(h_t \mid x_{1:t})\, dh_t$

(ii) $p(h_{t+1} \mid x_{1:t+1}) = 1/k_{t+1}\, p(x_{t+1} \mid h_{t+1})\, p(h_{t+1} \mid x_{1:t})$

Figure 8. Depiction of the posterior updating process from *t* to *t+1* in the Bayesian filtering algorithm.

Bayesian filtering algorithm.

Initialization. The initial posterior (i.e. the prior) is $p(h_0)$, which is the initial distribution of the hidden variable of the generative process.

Updating. Following each observation x_{t+1}, the distribution is updated from $p(h_t \mid x_{1:t})$ to $p(h_{t+1} \mid x1:_{t+1})$ in two steps.

(i) Go from $p(h_t \mid x_{1:t})$ to $p(h_{t+1} \mid x_{1:t})$ by taking into account the dynamics of the generative process. This calculation is performed with the equation:

$$p(h_{t+1} \mid x_{1:t}) = \int p(h_{t+1} \mid h_t)\, p(h_t \mid x_{1:t})\, dh_t \quad \textbf{[4]}$$

(ii) Add the new observation x_{t+1}. The new posterior is calculated with the equation:

$$p(h_{t+1} \mid x_{1:t+1}) = 1/k_{t+1}\, p(x_{t+1} \mid h_{t+1})\, p(h_{t+1} \mid x_{1:t}) \quad \textbf{[5]}$$

where $k_{t+1} = p(x_{t+1} \mid x_{1:t})$, is a normalisation constant that can be taken into account in the computation implicitly, by dividing the unnormalized distribution by its sum across the possible values of h_{t+1}, so that the resulting distribution sums to 1.

The equations **[4]** and **[5]** hold due to the structure of the generative process, shown in **Figure 1**, which has two key conditional independence properties: (a) h_{t+1} is independent of $x_{1:t}$ given h_t, which allows for equation **[4]** to hold true, and (b) x_{t+1} is independent of $x_{1:t}$ given h_{t+1}, which allows for equation **[5]** to hold true. The proofs are given below.

Proof of [4].

$$p(h_{t+1} \mid x_{1:t}) = \int p(h_{t+1}, h_t \mid x_{1:t})\, dh_t \qquad \text{(sum rule)}$$
$$p(h_{t+1} \mid x_{1:t}) = \int p(h_{t+1} \mid h_t, x_{1:t})\, p(h_t \mid x_{1:t})\, dh_t \qquad \text{(chain rule)}$$
$$p(h_{t+1} \mid x_{1:t}) = \int p(h_{t+1} \mid h_t)\, p(h_t \mid x_{1:t})\, dh_t \qquad \text{(property (a))}$$

Proof of [5].

$$p(h_{t+1} \mid x_{1:t+1}) = p(x_{t+1} \mid h_{t+1}, x_{1:t})\, p(h_{t+1} \mid x_{1:t}) / p(x_{t+1} \mid x_{1:t}) \qquad \text{(Bayes' rule)}$$
$$p(h_{t+1} \mid x_{1:t+1}) = p(x_{t+1} \mid h_{t+1})\, p(h_{t+1} \mid x_{1:t}) / p(x_{t+1} \mid x_{1:t}) \qquad \text{(property (b))}$$

An example of computed posteriors for a sequence of a probability learning task is illustrated in **Figure 9**.

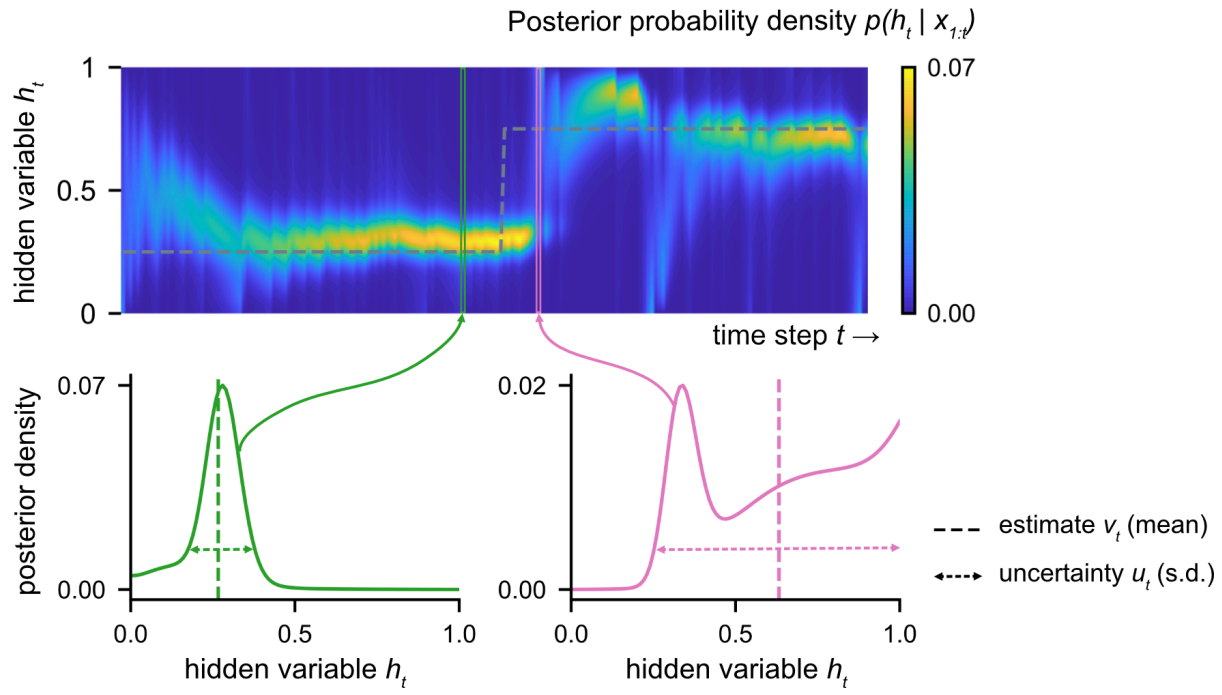


Figure 9. Example of posteriors calculated over a sequence of observations in a probability learning context. A sequence of binary observations was generated according to the hidden probability value shown by the grey dotted line on the top plot, with a change point in the middle. The top plot shows the (optimal) posterior probability density for each time step of the sequence as a colormap (bright yellow and dark blue indicate high and low probability density, respectively). The two line plots at the bottom show in more detail the posterior for the two time steps outlined in red and pink above.

3.3. Normative properties governing the adaptations of the learning rate

Several principles can be derived from the optimal solution presented above. Out of these principles, the focus of my thesis is mainly on the effect of uncertainty, which, because it varies dynamically with each observation, is a driver of the dynamic adjustments of the learning rate (level 2 of adaptive learning described above).

Effect of volatility: *The higher the volatility of the environment, the larger the updates.*

This effect is reflected in the term $p(h_{t+1} \mid h_t)$ of equation **[4]**. When the volatility is zero, $p(h_{t+1} \mid h_t)$ is zero everywhere except for $h_{t+1}=h_t$, which results in equation **[4]** inducing no update to the posterior. The higher the volatility, the larger $p(h_{t+1} \mid h_t)$ becomes for values of h_{t+1} unequal to h_t, which leads to a larger updating of the posterior.

Effect of stochasticity: *The higher the stochasticity of the environment, the smaller the updates.*

This effect is reflected in the likelihood term $p(x_{t+1} \mid h_{t+1})$ of equation **[5]**. Stochasticity is maximal when the likelihood function is uniform, meaning that all values of h_{t+1} lead to an equal likelihood value. In that case, equation **[5]** does not produce any update to the posterior. The lower the stochasticity, the more the likelihood term favours certain values of h_{t+1} over others (those values that are more consistent with the observation received x_{t+1}), which leads to a larger updating of the posterior by equation **[5]**.

Effect of uncertainty: *The higher the uncertainty associated with the estimate, the larger the update (and thus the learning rate) triggered by the new observation*.

Uncertainty is present in the posterior $p(h_t|x_{1:t})$ being updated. In statistical terms, uncertainty corresponds to the dispersion of the posterior distribution: the more dispersed the distribution, the more probable it is that the true value of h_t is far from the (point) estimate derived from the posterior, resulting in greater uncertainty associated with this estimate. One measure of dispersion is the standard deviation. Unless stated otherwise, I will use this measure to quantify the normative uncertainty (unless otherwise stated:

$$u_{t+1} = SD[h_t \mid x_{1:t}] \quad \textbf{[6]}$$

Remark: this time indexing is used so that α_{t+1} is regulated by the variable u_{t+1}, rather than u_t.

Normatively, uncertainty affects the updating during the multiplication between $p(h_{t+1} \mid x_{1:t})$ and the likelihood function $p(x_{t+1} \mid h_{t+1})$ in equation **[5]**. A graphical illustration of this effect is provided in **Figure 10**. Here is a simplified textual explanation. When

uncertainty increases, $p(h_{t+1} \mid x_{1:t})$ is more spread out and its probability density becomes larger far from its centre of mass. Let's assume that the centre of mass of $p(x_{t+1} \mid h_{t+1})$ is shifted to the right compared to $p(h_{t+1} \mid x_{1:t})$. Once multiplied with $p(x_{t+1} \mid h_{t+1})$, the probability density of the posterior will increase towards the right and decrease towards the left, and these increases and decreases in probability density will be all the higher than the original probability density was high, that is, that the uncertainty was high. Higher uncertainty therefore produces a larger shift of the posterior towards the right, a larger shift in posterior mean, a larger estimate update (the numerator of equation **[1]**), and thus a larger learning rate (having kept the denominator of equation **[1]** constant).

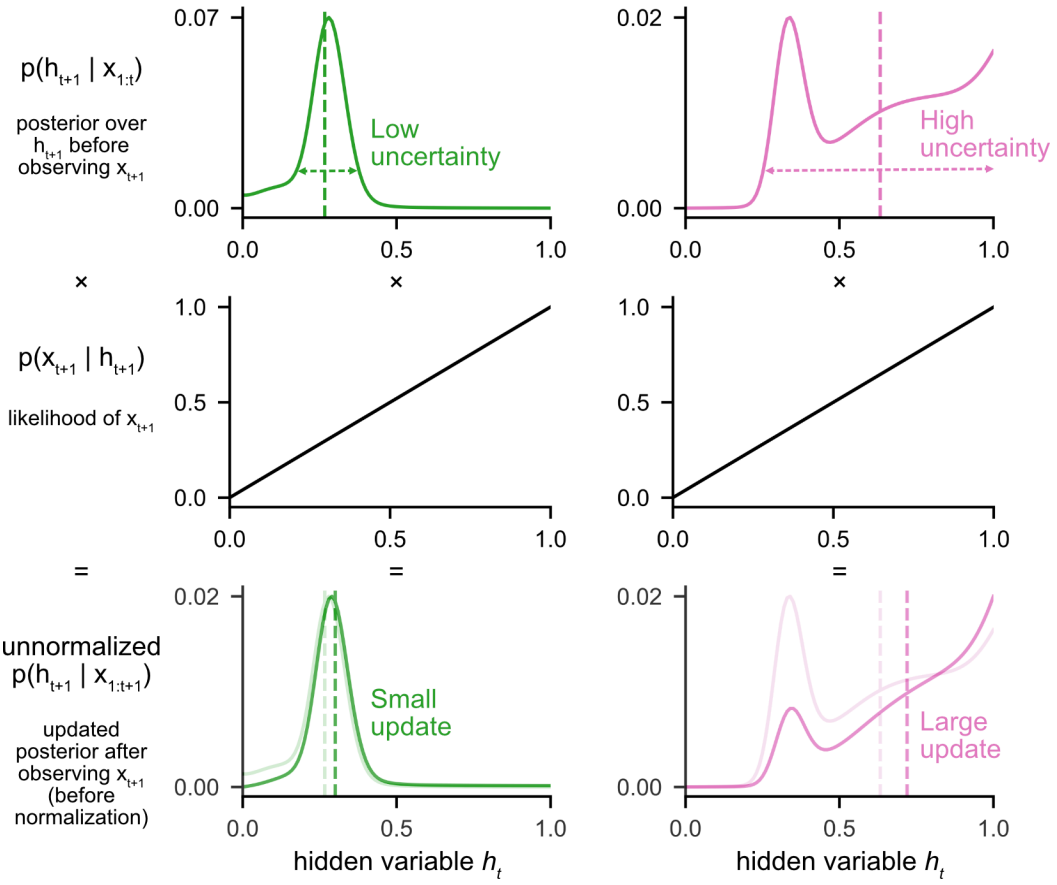


Figure 10. Example illustrating the effect of uncertainty on the update during the multiplication with the likelihood function in equation [5]. The posteriors at the top were extracted from the two time steps of the example shown in Figure 9. The likelihood function in the middle is that of a probability learning context for x_{t+1}=1.

3.4. Computational models of learning

Numerous models have been proposed in the literature on learning. I focus here on those that have been used in the behavioural and brain sciences as models of learning in humans and non-human animals. These models are often inspired by algorithms from computer science, engineering, and machine learning. I will list and detail these models below.

As we will see, these models have been designed "by hand" (handcrafted) by researchers, following a traditional approach that has the merit of producing easily explainable and interpretable models. By exploring these models, we will have a better understanding of the different approaches adopted and also the limitations of these approaches (which will be summarised in the next section). To give an overview, pure normative approaches produce optimal principles-based models but are commonly criticised for their possible too-high complexity; therefore, some researchers try to produce models of lesser complexity by using simplifications which unfortunately have consequences that are not well determined, while others restrict the problem to a particular type of generative process, which limits the applicability of the model, or use a model that assumes a particular generative process and apply the model to another generative process, which produces unpredictable behaviour.

Delta-rule / Rescorla-Wagner model (Rescorla & Wagner, 1972)
Normative models (Adams & MacKay, 2007; Behrens et al., 2007; Chen, 2003; Meyniel et al., 2015)
Reduced normative model for magnitude learning (Nassar et al., 2010, 2012)
Approximation of normative models by sampling: Particle Filters (Gordon et al., 1993; Chen, 2003)
Kalman Filter (Kalman, 1960)
Kalman filter-based model with estimation of volatility and stochasticity (Piray & Daw, 2021)
Hierarchical Gaussian Filter (C. Mathys et al., 2011; C. D. Mathys et al., 2014)
Adaptive mixture of delta-rules (Wilson et al., 2013)
Proportional-Integral-Derivative (PID) controller (Ritz et al., 2018)
Metaplastic synapses guided by a change detection system (Iigaya, 2016)

List 1. Computational models of learning proposed in the existing literature. The models are described below.

Delta-rule / Rescorla-Wagner model (Rescorla & Wagner, 1972).

The delta-rule has been introduced earlier in this manuscript, see equation **[2]**. It is the simplest learning model among the models discussed here. It serves as a baseline for the definition of adaptive learning: by construction, this model has a fixed learning rate and its learning is therefore non-adaptive. Note that, as mentioned in **Box 2**, the delta-rule is often embedded as a component in larger models, including all reinforcement learning models based on temporal difference (TD) learning such as Q-learning and SARSA (Schultz et al., 1997; Sutton & Barto, 2018).

Normative models (Adams & MacKay, 2007; Behrens et al., 2007; Chen, 2003; Meyniel et al., 2015)

Normative models realise the optimal solution to the learning problem for the generative process of the given task. In section 3.2 above, I described a general way to compute the optimal solution using Bayesian filtering (Chen, 2003). Other ways to compute the optimal solution have been proposed for specific classes of generative processes. In particular, for those with change points, Adams and MacKay (Adams & MacKay, 2007) proposed another algorithm that, in addition to providing estimates of the hidden quantity, provides an estimate of when the last change point occurred (it computes the posterior probability distribution over the run length, which is the number of observations since the last change point). (Meyniel et al., 2015) also proposed another algorithm to estimate change points based on sampling. This is interesting because some laboratory tasks require subjects to detect change points (Gallistel et al., 2014). However, these change-point estimation algorithms are significantly computationally more costly than Bayesian filtering.

One note regarding when a model is considered 'normative' in this thesis. Generally, ‘normative’ means that the model realises the optimal solution. Sometimes, such a model can be used with some assumed parameter values of the generative process which may differ from the true ones, such as the level of volatility or the level of stochasticity. This is typically the case when these parameters are fitted to the subject’s behaviour rather than set to the true generative ones. Even though the model no longer performs optimally when the model and the generative parameter values differ, I still refer to the model as a normative model because it *can* perform

optimally (with the right settings) and performs the same computational process as an optimal model. However, I no longer consider a model to be normative when it assumes a different *structure* of the generative process compared to the true one (for example, when the model assumes that the dynamics follow a random walk process while in reality they follow a change point process), as in that case it cannot perform optimally and may involve a very different computational process from that of any optimal model.

Reduced normative model for magnitude learning (Nassar et al., 2010, 2012).

Nassar et al. proposed a reduced (i.e. simplified and approximated) version of the normative model for the specific case of magnitude learning, in which the reduced model achieves a performance to that of the full model (Nassar et al., 2010). The approximations include that the reduced model only maintains the first two moments (the mean and the variance) of the posterior distribution, instead of maintaining the complete distribution, and that it computes them according to a formula that considers only a single expected run length (rather than all possible run lengths). These approximations greatly reduce computational complexity.

One advantage of this reduced model is that it explicitly calculates the learning rate, using an analytical formula which can serve as a simplified explanation for how the learning rate should be adjusted in the case of magnitude learning. The formula is: $\alpha_t = \tau_t + (1-\tau_t)\ \Omega_t$, where Ω_t is the change-point probability and τ_t is the relative uncertainty (Nassar et al., 2012).

While the performance of the reduced model is nearly optimal for magnitude learning, it may fall short in other learning contexts. Among others, one reason why it is likely to fall short is that by approximating the posterior by the first two moments, it loses any bimodal or asymmetric features that the normative posterior may have. And indeed, in certain cases, such as in probability learning, the normative posterior distribution can have highly asymmetric and bimodal shapes, and these features greatly influence the subsequent updates and thus performance of the model : see the pink distribution in **Figure 9** and **10**. (Also note that asymmetry in the posterior leads to an asymmetric update between positive and negative prediction errors, at equal absolute prediction error, which cannot be reproduced by the reduced model.)

Approximation of normative models by sampling: Particle Filters (Brown & Steyvers, 2009; Chen, 2003; Gordon et al., 1993; Prat-Carrabin et al., 2021).

The optimal solution involves calculating integrals that are often difficult to compute exactly (see for example equation **[4]**). Such an integral can be treated as an expectation, which can be approximated by sampling from the underlying distribution and computing the sample mean corresponding to that expectation. This idea has given rise to a family of methods and algorithms known as sampling-based approximation. The *particle filter* (Chen, 2003; Gordon et al., 1993) is one such sampling-based method that is specifically designed to approximate Bayesian filtering, thus providing a model for learning.

In the particle filter, the posterior distribution is approximated by a set of samples (called "particles") and a set of weights associated to those samples, which represent their respective posterior probabilities. The algorithm provides a guarantee: as the number of samples tends to infinity, it produces the optimal solution. The number of samples controls the degree of approximation, and can be treated as a model parameter.

Brown and Steyvers (Brown & Steyvers, 2009) and Prat-Carrabin et al. (Prat-Carrabin et al., 2021) have applied the particle filter to the modelling of human behaviour in magnitude learning tasks, fitting the number of samples in the model to the subject's behaviour. This produced a better fit to behaviour compared to the other models they tested.

Kalman Filter (Kalman, 1960).

Owing to its relative simplicity and its general usefulness in a wide range of engineering applications, the Kalman filter has been used also in neuroscience, to model for example the integration of visual information in visuomotor tasks (Baddeley et al., 2003; de Xivry et al., 2013), and as a basis for other neuroscience learning models (see the next two models presented below).

The Kalman filter is an algorithm designed to solve the learning problem in the specific case where the generative process is assumed to be the Gaussian random walk described in section 3.1, corresponding to the context of magnitude learning

with random walk dynamics. For this specific generative process, the optimal solution can be computed much more easily than in the general case. Because the posterior in that case is a Gaussian distribution, only the first two moments of the posterior distribution need to be maintained: the mean v_t and variance w_t of the posterior. Furthermore, these can be calculated using a simple set of (tractable) update equations (Kalman, 1960), which are provided below:

$$v_t = v_{t-1} + \alpha_t (x_t - v_{t-1}) \quad \textbf{[7]}$$

$$w_t = (1-\alpha_t) (w_{t-1} + \sigma_h^2) \quad \textbf{[8]}$$

where α_t is the apparent learning rate, which is calculated by the equation:

$$\alpha_t = (w_{t-1} + \sigma_h^2) / (w_{t-1} + \sigma_h^2 + \sigma_x^2) \quad \textbf{[9]}$$

Note that in the Kalman filter, the learning rate (α_t) and the uncertainty (w_t) do not depend on the observed sequence ($x_{1:t}$), but only on the number of observations received from the start of the sequence (t). In that sense, the Kalman filter does not perform level 2 adaptive learning. Also note that, as the coefficient ($1-\alpha_t$) in equation **[8]** is less than 1, the uncertainty w_t tends to decrease over time, and so does the learning rate α_t (since α_t is an increasing function of w_t according to equation **[9]**). In an environment with change points, this monotonous decrease of the learning rate would not be adaptive, as the adaptive solution in that case is characterised by dynamic increases in the learning rate in response to change points (see section 2.2 and chapter 2).

Kalman filter-based model with estimation of volatility and stochasticity (Piray & Daw, 2021)

The model by Piray and Daw extends the equations of the Kalman filter presented above and promotes the volatility σ_h^2 and stochasticity σ_x^2 parameters of the model, originally fixed, into dynamic variables. As presented by Piray and Daw, this extension allows the model to estimate volatility and stochasticity.

In the context of my thesis, what I find particularly interesting in the results that Piray and Daw presented with this model, is that when it is run on sequences generated by

a process with change point dynamics (which aren't the dynamics that the model assumes, but the model can still be run on those sequences as if they had been generated by a process with random walk dynamics), the model exhibits dynamic increases of the learning rate in response to change points (which are absent from the original Kalman filter, as mentioned above) (Piray & Daw, 2020, 2021). These dynamic increases in the Piray and Daw model reflect the increases in the estimated volatility variable of the model: when a change point occurs, the Piray and Daw model estimates that the assumed volatility of the random walk has increased (even when the true generative volatility, i.e. the average rate of change points, is held fixed).

This observation illustrates the fact that a normative learning model which assumes random walk dynamics but with variable volatility (like the Piray and Daw model), and a normative learning model which change point dynamics with fixed volatility (like the one used in (Foucault & Meyniel, 2021, 2023; Meyniel et al., 2015; Nassar et al., 2012)), both predict dynamic increases of the learning rate when exposed to the same sequences generated by a process with change point dynamics. A qualitatively similar behavioural signature can thus be exhibited by these two types of model, which provide a different interpretation of the underlying learning process that gives rise to this behaviour.

Hierarchical Gaussian Filter (C. Mathys et al., 2011; C. D. Mathys et al., 2014)

Like the Kalman filter and the Piray and Daw model, the Hierarchical Gaussian Filter (HGF) assumes a generative process in which the hidden variable follows a Gaussian random walk process. The generative process assumed by the HGF has multiple hidden variables that evolve in parallel and are organised hierarchically. In the case where there is only one level in the hierarchy, the model is equivalent to the Kalman filter. In the general case, the hidden variable at level $i+1$, $h^{(i+1)}$, and the hidden variable at level i, $h^{(i)}$, are coupled by the fact that the noise variance of the random walk of $h^{(i)}$ is equal to a positive (exponential) function f_i of $h^{(i+1)}$, i.e.:

$$h_{t+1}^{(i+1)} \sim N(h_t^{(i)}, f_i(h_{t+1}^{(i+1)}))$$

As computing the exact optimal solution for the generative process assumed by the HGF is difficult, the HGF model approximates the optimal solution using a variational inference method. This method approximates the posterior distribution over the hidden variable at each level as a Gaussian distribution.

In experimental studies, the HGF is often used with two hierarchical levels, which introduces a dynamic volatility variable similar to that in the Piray and Daw model (Piray & Daw, 2020). Note that, like the Piray and Daw model, the HGF model has been applied in cases where the sequences were generated by a process different from that assumed by the HGF, such as one with change point dynamics. In these use cases, the quality of the solution provided by the HGF remains unclear: in the results presented by (Piray & Daw, 2020), the HGF exhibited learning rate adjustments that go in the opposite direction of what is expected normatively (a decrease instead of an increase in the learning rate following a change point).

Adaptive mixture of delta-rules (Wilson et al., 2013).

Wilson et al. have proposed a learning model that relies on an adaptive mixture of delta-rules which corresponds to an approximation of the full Bayesian model with change-point inference presented in (Adams & MacKay, 2007). In this model, the estimate of the hidden variable (v_t) is computed as a weighted sum (i.e., a mixture) of the estimates computed in parallel by different delta-rules with different learning rates. It is an *adaptive* mixture because the weights assigned to each delta-rule are adjusted at each time step.

This model is related to the full Bayesian model by Adams and MacKay in the following way. The full Bayesian model computes a posterior distribution over all possible run lengths. In the Wilson et al. model, this posterior is approximated using a subset of possible run lengths, r_1, …, r_k. The estimation of the hidden variable assuming that the run length is equal to r_i is performed by the i-th delta-rule of the Wilson et al. model. The weight assigned to delta-rule i on a given time step is calculated based on the posterior probability of the corresponding run length, $p(r_i \mid x_{1:t})$.

Proportional-Integral-Derivative (PID) controller (Ritz et al., 2018).

The proportional-integral-derivative (PID) controller was developed in a branch of engineering called control theory, which studies the behaviour of dynamical systems. The learning process can be seen as a control system in which the output (the control signal) corresponds to the update of the estimate of the hidden variable ($\Delta v_t = v_t - v_{t-1}$), and the error signal (the feedback signal which regulates the control system) corresponds to the prediction error ($\delta_t = x_t - v_{t-1}$).

The PID controller is a simple model where the output signal is computed as a function of the error signal by adding three terms: the proportional term, which is proportional to the current value of the error signal, the integral term, which is proportional to the integral over time of the past values of the error signal, and the derivative term, which is proportional to the current value of the derivative of the error signal. In the discrete-time case, the integral and derivative terms become the sum of the past errors and the difference in the error between the last and current time step. This gives the following update equation:

$$\Delta v_t = K_P \delta_t + K_I \sum_{i=1}^{t} \delta_i + K_D (\delta_t - \delta_{t-1})$$

Ritz et al. (Ritz et al., 2018) applied this model to the modelling of human behaviour in a magnitude learning task. Their version of the PID controller was modified to introduce a leak factor in the integral term. They reported that their PID model provided a better fit to behaviour in their task compared to the delta-rule and the Kalman filter.

Metaplastic synapses guided by a change detection system (Iigaya, 2016).

Iigaya has studied the modelling of adaptive learning through a neural network in which the synapses are metaplastic, meaning that the rates of synaptic plasticity (i.e. the extent to which the synaptic weights change from one time step to the next) can change (Iigaya, 2016). In the Iigaya model, the modification of the rates of synaptic plasticity is regulated by a change detection system (referred to as surprise detection system in the original paper). It is through this system that the system can exhibit an adaptive apparent learning rate. The change detection system is implemented by a

mathematical algorithm whose biological feasibility remains unknown (this has been left to future studies, as per the author). Thus, the biological feasibility of the Ligaya model as a whole remains unknown.

If you made it to this sentence without skipping, congratulations and thank you for your careful reading.

3.5. Open questions: Feasibility and mechanisms of adaptive learning in the brain and neural networks

Overall, the models listed in the previous section have an undetermined capacity for generalisation or a computational complexity that may be too high. Even more critically, their feasibility and how they could be implemented in the brain remain unknown (except for the delta-rule, which does not possess the desired adaptive properties). Indeed, these models are described by equations and algorithms that are not constrained by the brain. They are also limited by the researchers' design capabilities (Saxe et al., 2021). How can adaptive learning be realised in the brain?

In the study presented in Chapter 3, I overcome the above limitations by adopting a new approach, which partially automates the design of the model and allows a large class of possible models to be explored, while being constrained by feasibility in the brain. I rely on recurrent neural networks (RNNs) (Goodfellow et al., 2016), a class of artificial neural network models whose feasibility in the brain is established because their basic building blocks are all biologically realisable (Douglas & Martin, 2007; Hunt & Hayden, 2017; R. C. O'Reilly & Frank, 2006), and which have led to advances in several neuroscience domains (Chaisangmongkon et al., 2017; Dubreuil et al., 2022; Mante et al., 2013; Sheahan et al., 2021; Wang et al., 2018). Optimization and machine learning algorithms are used to automate the assembly of the defined building blocks and the adjustment of the network's parameters to solve the given learning problem, thus producing a biologically feasible model of learning. Through this method, I've explored numerous possible architectures and sizes of RNNs to study the feasibility and minimal sufficient mechanisms to achieve the desired adaptive properties in learning. The specific questions and investigation methods of this study will be introduced in section 5 of this introduction and fully detailed in the article in Chapter 3.

4. Neural bases of adaptive learning

In this section, I review the findings about the neural bases of different components of learning. These components are involved in most probabilistic learning models: the surprise (or prediction error), the update, and the uncertainty (the latter being critical for adaptive learning, as explained in section 3). They can be articulated around a schematic model of the neural learning process hypothesised in many studies investigating the neural basis of learning, represented in **Figure 11**. However, not all components of learning have a clearly established neural basis; this includes probabilities (see section 4.4 and Chapter 4).

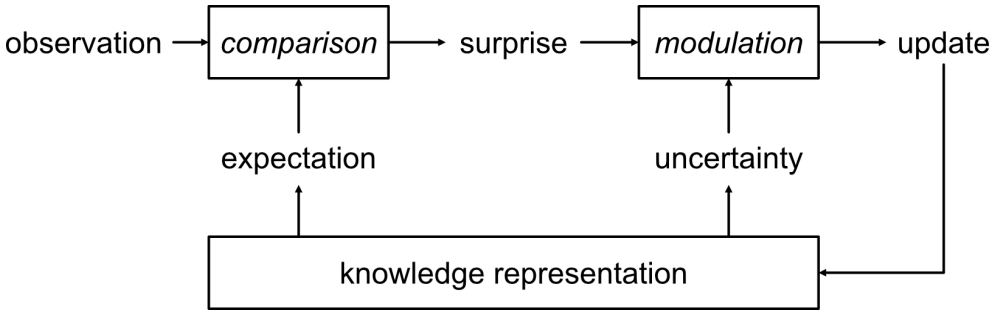


Figure 11. Schematic model of the hypothetical neural learning process. The received observation produces a surprise commensurate with its deviation from what was expected based on prior knowledge. The surprise, in turn, leads to an update of the knowledge represented in the brain. The uncertainty modulates the updating of knowledge by surprise.

4.1. Surprise signals emerge in the brain in response to an observation

The ***surprise*** or ***prediction error*** is a quantity that is central in many theories of brain function. These theories include *predictive coding*, the *free energy principle*, *reinforcement learning* and *optimal control* theories, as well as other theories subscribing to a Bayesian, Helmholtzian view of the brain as an *inference* machine (Daw & Doya, 2006; Dayan et al., 1995; Friston, 2010; Knill & Pouget, 2004; Rao & Ballard, 1999; Rescorla & Wagner, 1972; Schultz et al., 1997; Sutton & Barto, 2018). Surprise or prediction error is involved in these theories as a quantity that the brain seeks to minimise or maximise: predictive coding minimises prediction error; the free

energy principle minimises free energy which has two terms, one of which is equal to surprise; optimal control and reinforcement learning maximise value which is often estimated using reward prediction errors, and this maximisation can often also be written as a maximisation of likelihood and a minimisation of surprise. In general, the optimised quantity is often proportional to, or a function of, or contains, surprise or prediction error. Surprise and prediction error both measure the deviation between the just-received observation and the expectation (= prediction) formed on the basis of prior knowledge. In the strict sense, one definition of surprise, according to Shannon, is the negative logarithm of the expected probability of the observation occurring, and one definition of prediction error is the difference between the value of the observation and that of the expectation. In this section, I adopt a broader definition of "*surprise*" that encompasses the two previous definitions (*Shannon surprise* and *prediction error*).

Some of the earliest discovered neural correlates of surprise are the electrophysiological correlates observed in oddball and oddball-like paradigms, in which a surprising stimulus (called oddball) can occur within a sequence of stimuli (Bendixen et al., 2007; Horvath et al., 2001; Lieder et al., 2013; Mars et al., 2008; Squires et al., 1976). In response to the oddball stimulus, evoked potentials are observed including the mismatch negativity (MMN) and the P300.

In probabilistic learning paradigms, within the brain, surprise signals are observed first in the sensory areas associated with the modality in which the observation is perceived: visual (Bounmy et al., 2023; McGuire et al., 2014; Meyniel & Dehaene, 2017), or auditory (Meyniel & Dehaene, 2017; Ulanovsky et al., 2004). In the case where observations are rewards, surprise signals, which are more often referred to as reward prediction error (RPE) signals in that case, are also observed in a dopaminergic circuit that includes the mesolimbic pathway and the nigrostriatal pathway, involving many subcortical and cortical structures: the ventral tegmental area (VTA), substantia nigra compacta (SNc), ventral striatum (including nucleus accumbens) and other basal ganglia nuclei, the amygdala, the ventromedial prefrontal cortex, and when punishments are included, the anterior insula and dorsal striatum (Daw et al., 2006; Garrison et al., 2013; Palminteri et al., 2012; Schultz, 2022; Schultz et al., 1997). However, in these regions the neural correlates of

surprise are not domain-general: they are bound to the sensory modality or the rewarding nature of the observation.

Across sensory modalities and independently of rewards, surprise signals are observed in higher-level regions that include parietal, prefrontal, and anterior cingulate cortex areas (Bounmy et al., 2023; Kouider et al., 2015; McGuire et al., 2014; Meyniel & Dehaene, 2017; J. X. O'Reilly et al., 2013; Strange et al., 2005). These regions encode surprise across multiple experimental paradigms and multiple learning contexts (magnitude learning, probability learning). In these regions, the neural coding of surprise therefore seems to be domain-general, and surprise is encoded monotonically, and most often positively: the higher the surprise, the higher the neural activity.

4.2. Surprise signals trigger update signals

Knowledge updates are typically triggered by a surprise signal, and can be distinguished from surprise in two ways: normatively, the knowledge update is modulated by other factors, and behaviorally, knowledge updates can be measured from the subject's responses.

Knowledge updates have repercussions in many cortical regions, predominantly fronto-parietal, as observed in fMRI (Meyniel & Dehaene, 2017; J. X. O'Reilly et al., 2013) and in EEG/MEG (Fischer & Ullsperger, 2013; Jepma et al., 2016; Meyniel, 2020). Greater updating has also been related to increased functional connectivity between the fronto-parietal network and other functional systems that include regions in the dorsolateral and dorsomedial frontal cortex, the lateral and medial parietal cortex, and the anterior insula (Kao et al., 2020).

Updating is also reflected in pupil diameter (Filipowicz et al., 2020; Joshi et al., 2016; Meyniel, 2020; Nassar et al., 2012; J. X. O'Reilly et al., 2013). A study found that under constant surprise, pupil diameter correlates with the extent to which individuals updated their knowledge, as measured behaviorally (Filipowicz et al., 2020). A causal manipulation has also been conducted in the context of magnitude learning (Nassar et al., 2012). This manipulation involved playing an auditory stimulus increasing the subject's arousal level in a task-independent manner on certain trials. The increase in

arousal was reflected by an increase in pupil diameter. Concurrent with the effect on the pupil, it produced an effect on the subjects' learning rate, as measured behaviourally. The direction and magnitude of the effect on learning rates depended on the baseline level of pupil diameter.

The pupil diameter findings indicate that the locus coeruleus-norepinephrine (LC-NE) system (and possibly other neuromodulatory systems, e.g. cholinergic), could play a key role in updating. Indeed, pupil diameter is often considered to index LC activity because the LC controls pupil dilations via a sympathetic circuit in the spinal cord and the brainstem (Joshi & Gold, 2020). The LC is not the only structure controlling pupil diameter in general, but it is likely to be the one most contributing to non-luminance related changes in pupil size that occur during learning (Joshi & Gold, 2020). Anatomically, the connectivity of the LC puts it in a good position to integrate surprise signals and the other factors modulating the update, and to produce update signals widely distributed across the brain. The LC receives inputs from the cortex including the frontal and anterior cingulate regions, which convey surprise and uncertainty, and is the principal source of noradrenergic innervation of the entire cortex (as well as of the hippocampus, amygdala, cerebellum, and spinal cord) (Aston-Jones & Cohen, 2005). Functionally, a wide range of empirical findings show that the LC responds to the occurrence of a surprising stimulus, as measured not only through pupil dilations as mentioned earlier, but also directly using single-unit electrophysiological recordings of LC neurons (Aston-Jones & Bloom, 1981; Aston-Jones & Cohen, 2005; Joshi et al., 2016), and recently, using fMRI (Mazancieux et al., 2022).

4.3. Uncertainty signals may modulate the updates

Neural correlates of uncertainty (that is, estimation uncertainty, see **Note 1**) have been observed in the cortex in parietal, frontal, and cingulate regions across multiple learning contexts (magnitudes, probabilities) (Bounmy et al., 2023; McGuire et al., 2014; Meyniel & Dehaene, 2017). These regions are partly separate from, and partly overlap with surprise regions, as observed at the group level (McGuire et al., 2014; Meyniel & Dehaene, 2017), and at the subject level an overlap was also reported in the case of probabilities (Bounmy et al., 2023).

The correlates of uncertainty are also observed in MEG/EEG signals in the modulation of the response evoked by the observation (at equal levels of surprise, the amplitude of the response increases with uncertainty) and in the power of neural oscillations in the alpha-beta band (8-30 Hz) (Jepma et al., 2016; Meyniel, 2020). Furthermore, they are reflected in changes in pupil diameter (Meyniel, 2020; Nassar et al., 2012). Transient (also called “phasic”) and sustained (also called “tonic”) variations in pupil diameter reflect surprise and uncertainty, respectively (Joshi & Gold, 2020; Meyniel, 2020; Nassar et al., 2012).

Overall, these results combined with those from the above subsections relating pupil diameter to uncertainty, surprise, and updating suggest that: 1) uncertainty may modulate the activity of the LC-NE system and its responses to surprise, and 2) these modulations of LC-NE system activity may modulate the updates at the neural and behavioural level.

The hypothesis that the LC-NE system plays a regulatory role in learning is related to other theories on the functions of the LC-NE system and the mechanisms regulating learning. One theory that Yu and Dayan proposed is that NE signals unexpected uncertainty (which is strongly related to the uncertainty as defined in this thesis, since in Yu and Dayan's original model, it is equal to 1 - the maximum of the posterior probability distribution, and there is strong correlation between 1-max-p and SD), and that acetylcholine (ACh) signals expected uncertainty. They proposed that the uncertainty signal conveyed by NE interacts with the expected uncertainty signal conveyed by ACh to enable optimal learning in a stochastic and changing environment (Yu & Dayan, 2005). Another hypothesis (which follows in part from study 2 of this thesis, reported in Chapter 3) is that the modulation of learning rate by NE could occur through gain modulation, or gating. This hypothesis is supported biologically (by previous studies) and computationally (by study 2). I will return to this hypothesis in the general discussion.

4.4. Open question: Probability, a neural code yet to be discovered

One element that is missing from the discovered neural bases is probability. Probabilities are doubly involved when learning in a stochastic environment. Firstly, any stochastic environment is governed by probabilities which must be acquired,

used, and updated by the brain. Secondly, the learner's own knowledge (whether it concerns probabilities or other quantities) is normatively represented by probability distributions (see section 3). How are probabilities neurally represented in the brain?

Several proposals have been made regarding how a neural population could enable probabilistic computations. These proposals all attempt to explain how a neural population representing a scalar variable, such as the orientation of a stimulus, could also represent its probability density. Probabilistic population codes propose that under certain hypotheses, the probability density function can be decoded from neural activity (Beck et al., 2008; Ma et al., 2006), distributed distributional codes propose that neural activity encodes expectations calculated according to the probability density (Sahani & Dayan, 2003; Zemel et al., 1998), and neural sampling codes propose that neural activity successively represents samples (e.g. orientation values) drawn according to the probability density (Fiser et al., 2010; Hoyer & Hyvärinen, 2002). However, it is important to note that in all these proposals, the aim is to represent the probability density of a scalar variable, and not to represent the probability p of an event (such as the probability of occurrence of a given stimulus or outcome). None of these proposals attempt to explain how such a probability p may be encoded in neural activity (i.e., activity $a=f(p)$). Such a representation of probability is necessary though, at least in situations where individuals explicitly estimate one or more probabilities, possibly reason about them, and possibly report them or communicate them to others (for example to improve collective decision making), as required by task demands or because probabilities are useful to describe the world (Bahrami et al., 2010; Gluck et al., 2002).

This question is even more enigmatic given that neural correlates of probability have been searched for and, to my knowledge, using classical approaches, they have not yielded positive results (except for reward probability, which is a special case). Although negative results are rarely reported, inability to find linear correlations between probability and neural activity has been reported in (Bounmy et al., 2023; Lebreton et al., 2015). In Chapter 4, I examine this question using a new approach.

5. Objectives and research questions of the thesis articles

In Section 1, I presented the motivations and central issues in the study of adaptive learning in stochastic and dynamic environments. In Sections 2, 3, and 4, I presented what is known about the behavioural manifestations, the computational bases, and the neural bases of such learning, and highlighted at the end of each section a number of unanswered questions. Here, I present the work I conducted during my thesis to address these questions (as well as others, more specific to each study). I conducted three studies: a behavioural study, a theoretical study, and an fMRI study. Each study is the subject of an independent research article (published or forthcoming), included as a chapter in this thesis manuscript (Chapter 2, 3, and 4).

5.1. Article 1: Behavioural study of the dynamics of adaptive learning in magnitude and probability learning

Context and specificities of this study. Level 2 adaptive learning requires adapting the learning rate observation by observation (section 1.5). Normatively, the adaptations of the learning rate should be guided by uncertainty (section 3.3). To test these predictions empirically, it is necessary to study the dynamics of the learning rate at each observation, which remain unexplored in probability learning (section 2). Indeed, the investigation of these dynamics in humans has been limited to a narrow range of experimental paradigms, all of which concern magnitude learning and involve relatively low levels of stochasticity compared to those involved in probability learning. Study 1 of my thesis examines the dynamic adaptations in human learning and the factors that govern these dynamic adaptations, comparing probability learning and magnitude learning, and more generally testing the consistency of human behaviour to the normative model of learning presented in this thesis.

Among the three studies in my thesis, study 1 is the only one that, in addition to probability learning, studies magnitude learning, and examines the effect of surprise (or more precisely, change-point probability) on learning rates.

Questions investigated. Do humans dynamically adapt their learning rate during probability learning? How do their dynamic adjustments in probability learning compare to magnitude learning? What are the computational determinants that

govern their adjustments? What is the respective importance of these determinants for the adjustments in probability learning compared to magnitude learning? Do human adjustments and their determinants conform to normative theory?

Methods of investigation. To measure the subject's learning rate at each observation, I developed a new experimental paradigm that continuously tracks the updates made by the subject as observations occur. I developed two tasks using this paradigm, one for magnitude learning and the other for probability learning. I conducted a study where each of the two tasks was performed one after the other by the subjects. I compared the behaviour of the subjects with that predicted by the normative model to test their consistency. I identified and extracted from the normative model two determinants of the dynamic learning rate adjustments, uncertainty and change-point probability, to test their influence on the subjects' adjustments.

5.2. Article 2: Theoretical study of the feasibility and mechanisms of adaptive learning in neural networks

Context and specificities of this study. The feasibility and complexity of adaptive learning for the brain are quite mysterious. The numerous proposed models (listed in section 3.4) are specified in a computational form that does not reveal whether, how, and at what cost they could be implemented in the brain. Furthermore, the adaptive capabilities of these models (such as their ability to achieve the levels of adaptive learning listed in section 1.5, and their degree of conformance to the normative principles listed in section 3.3) are not always known and not often quantified in comparison to the optimal solution. Study 2 of my thesis seeks to clarify these questions using a class of artificial neural networks whose feasibility in the brain is established.

Study 2 examines a broader range of adaptive learning capabilities than those covered in the general introduction of my thesis. They are related to the presence of hierarchical latent structure in the environment, in addition to stochasticity and dynamics. These capabilities include, in addition to the dynamic adaptations of the learning rate, the ability to disentangle the hidden variables of the environment and use the relevant variable for the current context, as well as the ability to perform a

hierarchical transfer of learning from one variable to another when the variables are coupled by simultaneous change points.

Remark: I regret not mentioning in the article that I tested numerous state-of-the-art architectures more complex than the ones presented in the article, such as LSTM, and deep (multi-layer) RNNs (Goodfellow et al., 2016; Hochreiter & Schmidhuber, 1997). My goal was to determine the minimal sufficient architecture, and it turned out that after an extensive exploration of architectures, a simpler architecture (a one-layer GRU) was able to solve the problem just as well as the more complex architectures. That is why I did not mention them in the article. In hindsight, I acknowledge that mentioning them would have likely made the article more understandable and accessible to a broader audience.

Questions investigated. To what extent is the optimal (normative) solution to the problem of learning in a stochastic and dynamic environment feasible for the brain? Can biologically feasible artificial neural networks achieve quasi-optimal solutions? Which mechanisms are essential for achieving such solutions? What is the minimum sufficient network size to achieve such solutions? Qualitatively, do these networks exhibit the signatures of the normative Bayesian solution when confronted with increasingly complex environments (namely, the ability to dynamically adjust their learning rate, to disentangle the hidden variables, to hierarchically transfer their learning about one hidden variable to another)? Internally, how do the networks solve the problem? In particular, do they explicitly represent uncertainty (as the normative model does), and is that represented uncertainty epiphenomenal or crucial for their adaptive behaviour (as in the normative model)?

Methods of investigation.

- RNNs were trained to solve the learning problem in different environments with increasing complexity. In order to determine the best achievable performance for a given defined network architecture, the networks were trained as much as necessary and the training hyperparameters were determined by a thorough optimisation procedure.
- To determine which mechanisms are essential for achieving the desired performance and adaptive capabilities, mechanisms were systematically

added or removed one by one to define several network architectures, which were then trained and tested.

- A wide range of network sizes were tested for each architecture to determine their growth in performance with size and determine their minimum sufficient size needed to achieve the desired performance.
- The performance of the networks was systematically and quantitatively compared to the optimal solution and other classical models of learning. Each studied adaptive learning capability was examined in the networks through diagnostic behavioural tests, also in comparison to the optimal solution.
- The internal realisation of adaptive learning by the networks and their use of uncertainty were elucidated using an approach combining neuroscience methods (decoding, state-space analysis of the dynamics, causal perturbations) applied to the network's dynamic activities, techniques from linear algebra, and the unique advantages provided by the full access to the network's internal workings.

5.3. Article 3: fMRI study of the neural coding of probabilities during learning

Context and specificities of this study. As mentioned in section 4, the neural bases of probability, an essential component of probabilistic learning processes, remain largely unknown. Study 3 of my thesis seeks to determine them.

Questions investigated. How are probabilities represented in the brain? Is there, in mesoscopic-level brain activity, a neural representation that specifically represents probabilities rather than other confounding factors that correlate with probability in many studies? Which brain regions represent probabilities? How do their neural activities encode probability? What are the characteristics of the tuning curves for probability?

Methods of investigation. An experiment was conducted in which human brain activity was measured using fMRI during probability learning with change points (data reported in Bounmy et al. 2023; although this data has been reported in that study, the experiment was designed from the outset for the present study). To find the brain regions that encode probabilities, we developed a new approach based on versatile encoding models. These models can approximate arbitrary tuning curves and thus

allow us to test, from the measured fMRI signal, whether a region is selective for probability, without assuming a particular form of tuning curves. If there is selectivity, these models allow us to reconstruct the tuning curves of the selective voxels. A neural code for probabilities in the brain could thus be located, its tuning curves were reconstructed, and these were characterised using several characteristic measures. This characterisation was strengthened by comparing the results obtained for probability to those obtained for uncertainty (called confidence in this article), which was used as a control as its neural code was already relatively well known from previous studies. Numerous methodological controls were used to ensure that the determined neural code was specifically encoding probability rather than other confounding factors (stimuli, subject responses, surprise, expected rewards...), and to ensure that the results were reliable (encoding model generalisation to unseen data was tested by cross-validation, a stringent null hypothesis definition was adopted and controlled for in the test metric through null hypothesis simulations, the analysis pipeline was validated end-to-end through simulations in which experiments and fMRI signals were generated under different possible neural codes).

Chapter II: Article 1, Behavioural study (Foucault & Meyniel, 2023, submitted)

Preprint: https://www.biorxiv.org/content/10.1101/2023.08.18.553813

Two determinants of dynamic adaptive learning for magnitudes and probabilities

Cedric Foucault[1,2,*] and Florent Meyniel[1,*]

1. Cognitive Neuroimaging Unit, NeuroSpin (INSERM-CEA), University of Paris-Saclay, 91191 Gif-sur-Yvette, France

2. Sorbonne University, Doctoral College, F-75005 Paris, France.

* Corresponding authors: cedric.foucault@gmail.com and florent.meyniel@cea.fr

Keywords
Learning; probability; uncertainty; dynamic environment; continuous behavior.

Abstract

Humans face a dynamic world that requires them to constantly update their knowledge. Each observation should influence their knowledge to a varying degree depending on whether it arises from a stochastic fluctuation or an environmental change. Thus, humans should dynamically adapt their learning rate based on each observation. Although crucial for characterizing the learning process, these dynamic adjustments have only been investigated empirically in magnitude learning. Another important type of learning is probability learning. The latter differs from the former in that individual observations are much less informative and a single one is insufficient to distinguish environmental changes from stochasticity. Do humans dynamically adapt their learning rate for probabilities? What determinants drive their dynamic adjustments in magnitude and probability learning? To answer these questions, we measured the subjects' learning rate dynamics directly through real-time continuous reports during magnitude and probability learning. We found that subjects dynamically adapt their learning rate in both types of learning. After a change point, they increase their learning rate suddenly for magnitudes and prolongedly for probabilities. Their dynamics are driven differentially by two determinants: change-point probability, the main determinant for magnitudes, and prior uncertainty, the main determinant for probabilities. These results are fully in line with normative theory, both qualitatively and quantitatively. Overall, our findings demonstrate a remarkable human ability for dynamic adaptive learning under uncertainty, and guide studies of the neural mechanisms of learning, highlighting different determinants for magnitudes and probabilities.

Significance statement

In a dynamic world, we must constantly update our knowledge based on the observations we make. However, how much should we update our knowledge after each observation? Here, we have demonstrated two principles in humans that govern their updating and by which they are capable of dynamic adaptive learning. The first principle is that when they observe a highly surprising event indicating a likely change in the environment, humans reset their knowledge and perform one-shot learning. The second principle is that when their knowledge is more uncertain, humans update it more quickly. We further found that these two principles are differentially called upon in two key learning contexts that could be associated with different brain mechanisms: magnitude learning (which primarily requires adaptation to surprise, under the first principle) and probability learning (which primarily requires adaptation to uncertainty, under the second principle). Our findings advance understanding of the mechanisms of human learning, with implications for the brain and the development of adaptive machines.

Introduction

Humans live in a dynamic world that requires them to constantly update their knowledge as events unfold. For example, new construction works that cause significant delays in transportation should prompt users to revise their estimate of their typical commute time. This learning process is a fundamental part of our adaptive behavior, and a field of study that spans the sciences of behavior, brain, and machines (Rescorla & Wagner, 1972; Rosenblatt, 1961; Schultz et al., 1997; Sutton & Barto, 2018).

Learning is challenging because the environment is often not only dynamic, but also stochastic. When we take public transportation, for example, our commute time varies from day to day due to stochastic fluctuations. However, a delay can also be caused by an abrupt change, such as construction work, that may persist for several weeks. In this case, we face an ambiguity when a delay occurs: is it due to stochastic fluctuations (in which case we should not significantly change our estimated average commute time), or a genuine change point (in which case we should revise our estimate more drastically)? The challenge lies in determining the appropriate weight to give to the current observation (the observed delay) in our learning process.

A descriptive tool to quantify the weight given by the learner to an observation is the *apparent learning rate*, which we will simply call *learning rate* hereafter (Heilbron & Meyniel, 2019; Nassar et al., 2010). The learning rate measures the amount of update of the learned value (from v_{t-1} to v_t) induced by the observation x_t in proportion to its deviation from the previously learned value:

$$\alpha_t = (v_t - v_{t-1}) / (x_t - v_{t-1}) \quad [1]$$

This tool is useful for characterizing the learning process and distinguishing several levels of adaptive learning (Foucault & Meyniel, 2021; Soltani & Izquierdo, 2019). Level zero (non-adaptive) corresponds to a constant learning rate, as in the delta rule, a widely used learning model in psychology and neuroscience. This model accounts for the basic fact that the amount of update increases with the discrepancy between the expectation and the observation (also known as prediction error). It has led to numerous successes across various forms of learning, including associative learning and reinforcement learning in humans, animals, and artificial agents (O'Doherty et al., 2003; Rescorla & Wagner, 1972; Rosenblatt, 1961; Schultz et al., 1997; Sutton & Barto, 2018). Beyond level zero, adaptive learning processes involve an adaptation of the learning rate at some level. We distinguish two levels of adaptive learning: one is at the average level over a block of trials, and the other is more fine-grained, at the level of individual trials.

A first level of adaptive learning corresponds to adjusting the average learning rate to the global statistics of the environment. Two relevant statistics for adjusting the average learning rate are the average frequency of change points (also known as volatility) and the level of stochasticity in the environment (Soltani & Izquierdo, 2019). To illustrate, change points are more frequent in transportation systems where construction campaigns are more frequent, and commute times are less stochastic in transportation systems that are better organized. When change points are more frequent, the average learning rate should be higher to update the estimated value more quickly. Accordingly, previous studies have shown that humans use a higher average learning rate in a block of trials containing many change points compared to a block with no change points (Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019). When the environment is more stochastic, the average learning rate should be lower to stabilize the estimated value. Such an effect has also been observed in humans (Lee et al., 2020).

A second level of adaptive learning corresponds to dynamically adjusting the learning rate from one observation to the next depending on what is observed—we refer to it as *dynamic adaptive learning*. Such dynamic adjustments are particularly critical to learn effectively in a dynamic and stochastic environment, so as to increase the learning rate locally when a single change point is detected (Foucault & Meyniel, 2021; Nassar et al., 2010). Compared to the first level (different average learning rates between blocks of trials), less is known in humans about this second level (dynamic adjustments of the learning rate at the trial level).

Dynamic adaptive learning has been demonstrated in humans in the case of magnitude learning. Empirical evidence is available for several kinds of magnitudes including symbolic numbers (Nassar et al., 2012), positions (McGuire et al., 2014), and angular directions (Vaghi et al., 2017). However, these studies leave unexplored another common type of learning: probability learning. In the transportation example, this could be the probability that a traffic jam (or an incident in transit) occurs and causes us to be late.

The literature on probability learning is vast, from the learning of reward probabilities to stimulus occurrence to the validity of attentional cues, but it remains unknown whether humans dynamically adapt their learning of probabilities. Previous studies are not suitable to examine trial-level adjustments of the learning rate. Many of them are based on choices, from which only an average learning rate can be estimated, as a model parameter that is fitted across many choices and compared between blocks of trials (Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019; Cools et al., 2002).

From a learning perspective, magnitudes and probabilities present a fundamental difference: the amount of information provided by a single observation about the quantity to be learned is typically much lower for probabilities than for magnitudes (see Fig. S1 for common examples). This amount of information should in principle regulate the learning rate: the more information the observation provides about the quantity to be learned, the higher the learning rate should be. In particular, a single observation may be informative enough to detect a change point and thus immediately increase the learning rate. Conversely, when observations are less informative, change points are more ambiguous (difficult to distinguish from stochastic fluctuations), making dynamic adaptive learning all the more challenging.

In light of these differences, what has been demonstrated about human adaptive learning in the context of magnitudes may not generalize to probabilities. A comprehensive understanding of dynamic adaptive learning requires to study probability learning and to compare it directly to magnitude learning, in order to uncover the extent and determinants of dynamic adjustments in each case. Here, we address two main questions:

1) Do humans dynamically adapt their learning rate in probability learning (level 2, introduced above) or use a fixed learning rate (level 0)? What are their adjustment dynamics, and how do they compare to those of magnitude learning?

2) What are the computational determinants of the dynamic adjustments of learning rates?

To answer these questions, we conducted a study in which human subjects performed two learning tasks, a magnitude learning task and a probability learning task. We developed a new experimental paradigm to measure the subject's learning rate directly from their report at each observation. This measure of the learning rate serves to characterize the human learning process.

To identify the determinants of adjustments in learning rates, we adopted a normative approach (also known as rational analysis) (Lieder & Griffiths, 2020; Oaksford & Chater, 2007). We studied the optimal Bayesian solution to the learning problem posed and derived normative theoretical properties, which we then tested in subjects. We want to stress that we make no claims about the algorithm used by subjects to achieve these properties (we use the Bayesian solution as a means to identify these properties, not as a model of the algorithm).

We identified two computational determinants and we show that they play a quantitatively different role in the two types of learning. Overall, our results show a remarkable alignment of human behavior with normative behavior in dynamic adaptive learning, even in the difficult case of probability learning.

Results

Measuring the learning rate at each observation during magnitude or probability learning. To measure subjects' learning rate at each observation as they are learning, we developed a new experimental paradigm where subjects update their estimate in real time as they observe a sequence of stimuli occurring one at a time at regular intervals of 1.5s (Fig. 1). We obtained real time reports by continuous motion tracking: the estimation is done with a slider that follows the subject's motion, captured on a touchpad or a mouse (Fig. 1 A and B). The subject's estimate (v_t, obtained from the slider position) is thus collected at each observation (x_t, t indexes the observation). This allows us to calculate the learning rate for each observation using equation [1].

Note that the learning rate calculated in this way is directly obtained from the subject's estimates (Nassar et al., 2010), rather than derived from model fitting, as is often done (Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019). This learning rate is a behavioral measure that makes no assumptions about the computational process used by subjects for learning. The learning rate can thus be measured for all kinds of learning models and compared to subjects.

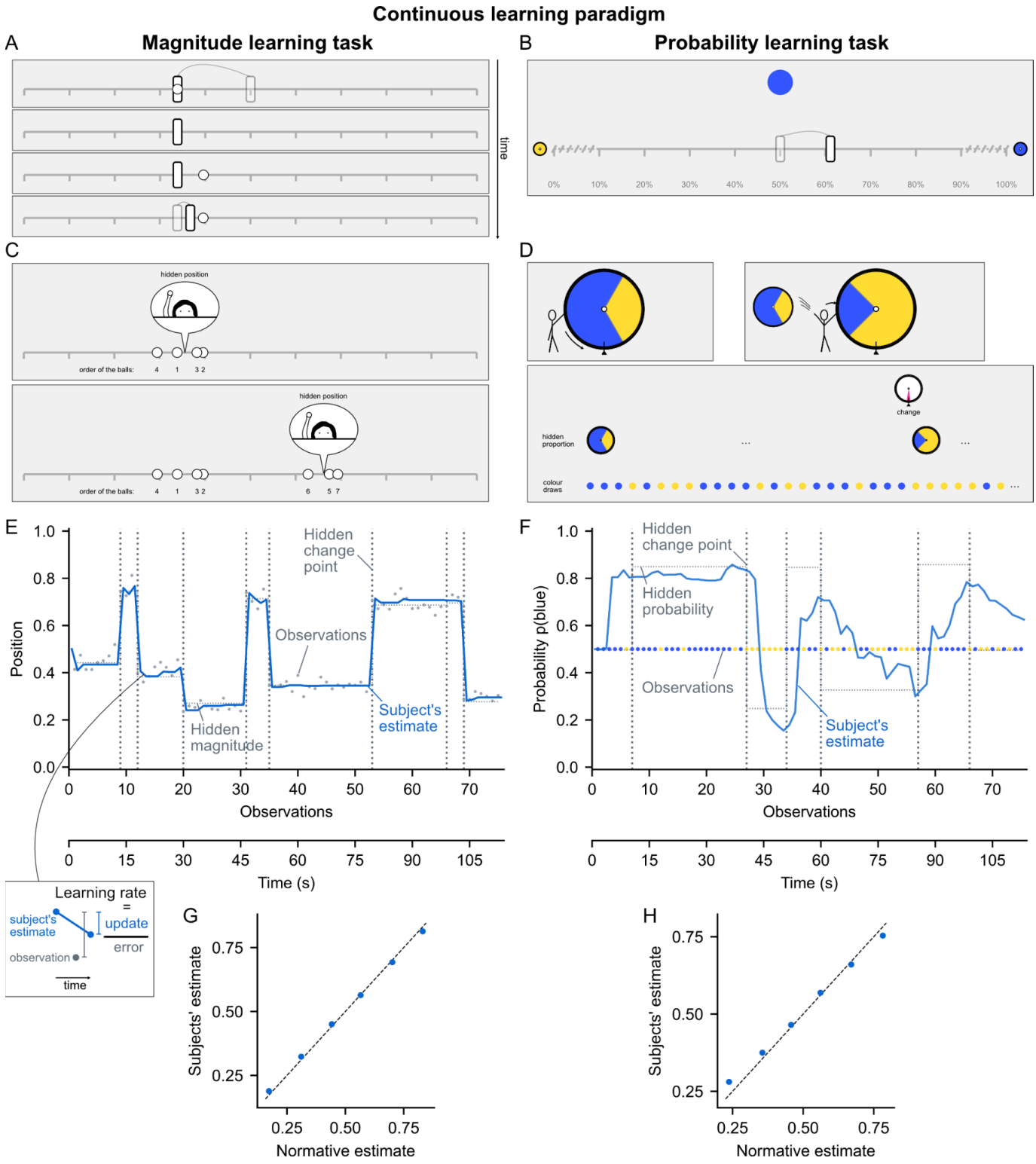


Fig. 1. Magnitude and probability learning tasks with estimates reported in real-time. (A and B) Screenshots of the two tasks. The subject must estimate a hidden quantity (magnitude or probability) based on the stimuli which appear one by one at regular intervals (1.5s). This quantity changes at hidden, unpredictable times called change points. The subject reports their current estimate by moving a slider in real time via motion tracking. For the magnitude task (A, frames show successive times), the hidden quantity is the mean horizontal position generating the stimuli (white circles). For the probability task (B), it is the probability that the circle appearing in the center is blue vs. yellow. Semi-transparent elements indicating slider movements were added here for explanation. (C and D) Screenshots of the task instructions. For the magnitude task (C), subjects were told that they should estimate the position of a hidden person throwing snowballs based on where the balls are landing (the white circles), and that the person could change position at random times without their knowledge. For the probability task (D), subjects were told they should estimate the proportion of blue and yellow on a hidden wheel used to draw the observed

colors, and that the wheel could be changed at random times without their knowledge. (E and F) Example of one session of the magnitude (E) and probability (F) task. Vertical dotted lines indicate the first observation after a change point. Thin dotted lines show the hidden quantity values. Dots represent observations (positions in (E), colors in (F)). The blue line shows the subject's estimate after each observation (0: left edge of the slider, 1: right edge). (G and H) Accuracy of subjects' estimates for each observation compared to normative (i.e. optimal) estimates in the magnitude (G) and probability (H) task. The data were binned in 6 equal quantiles of normative estimate and averaged within-subjects. Points and error bars (too small to be seen) show mean ± s.e.m. across subjects, dashed line is the identity line.

We developed two learning tasks using this paradigm: a magnitude learning task and a probability learning task. In order to compare their results, we adopted a similar design and task structure. In both tasks, the subject's goal is to estimate a hidden quantity (magnitude or probability) from the observations they receive, and to produce estimates as accurate as possible at each observation. During the observation sequence, the hidden quantity undergoes discrete changes at unpredictable random times called change points (not limited to reversals), which are also hidden (subjects are not told when a change point occurs). In the magnitude learning task, the hidden magnitude is the mean horizontal position generating the observed stimulus positions. Subjects were instructed that a hidden person was throwing snowballs and that their goal was to estimate the position of this hidden person based on where the snowballs are landing, knowing that the hidden person could change position from time to time without their knowledge (Fig. 1C). In the probability learning task, the hidden probability is the probability of the centrally-presented stimulus being blue (vs. yellow). Subjects were instructed that the colors were drawn by spinning a wheel filled with a certain proportion of blue and yellow and that their goal was to estimate this proportion, knowing that the wheel could be changed from time to time without being told when it changes (Fig. 1D). The observation sequences were generated, in the magnitude task, using the same parameters as (Nassar et al., 2012) for comparison (same or very similar parameters were also used in other magnitude learning studies (McGuire et al., 2014; Vaghi et al., 2017)). In the probability task, we used a probability of change point occurrence that made the task engaging and difficult but still tractable for subjects. (See Methods for more details.)

The human subjects (n=96) performed the two task one after the other, in a counterbalanced order across subjects (see an example session for each task Fig. 1 E and F). We evaluated subjects' estimate accuracy at the level of each observation by comparing their estimates with the normative estimates, i.e. the optimal estimates given the observations received. Although change points were frequent, subjects were able to produce accurate estimates (close to normative) of magnitudes and probabilities (Pearson r=0.96±0.01 and 0.80±0.01 mean±s.e.m., t_{95}>55.7, $p<10^{-73}$ in both cases, see Fig 1 G and H). (See Supplementary Text 1 for further details on the comparison between the normative estimates and the subjects' estimates.)

Subjects dynamically adapt their learning rate after a change point for probabilities, and differently than for magnitudes. As per question 1), we investigated subjects' ability to dynamically adapt their learning rate in response to change points, by calculating their learning rate with equation [1] and analyzing its dynamics around the change points. An optimal learner should increase their learning rate when they believe that a change point has occurred recently, to quickly update their estimate, and then decrease their learning rate to stabilize their estimate (see illustration of this behavior and how it differs from that of a learner using a fixed learning rate in Fig. 2C). Note that change points are more difficult to detect in the probability learning task because the amount of information provided by a single observation is much lower than for magnitude learning (Fig. S1). Despite this difficulty, subjects dynamically adapted their learning rate remarkably well, not only in the magnitude learning task, but crucially also in the probability learning task (Fig. 2 A and B).

In the magnitude learning task, the learning rate dynamics we found (Fig. 2A) replicate those found in previous studies of magnitude learning (McGuire et al., 2014; Nassar et al., 2012; Vaghi et al., 2017). The increase in learning rate occurs immediately and mainly at the first observation after a change point, with a learning rate suddenly close to 1. In the probability learning task, the learning rate dynamics show an increase that is smoother and more prolonged, remaining significantly higher at the eighth observation after a change point compared to before the change point (Fig. 2B). Importantly, these different dynamics are very similar to those prescribed by normative theory (see insets in Fig. 2; the Pearson correlation between subject's mean dynamics and the normative one is 0.99 in both tasks; the amplitudes of their dynamics are also similar to the normative ones, shown in further detail in Fig. S2). We also controlled for the frequency at which subjects updated: we observed similar, significant dynamic adaptations of the learning rate even after excluding from analysis all observations where subjects did not report an update (Fig. S4). (Note that subjects frequently updated their estimate in our study, see Fig. S3).

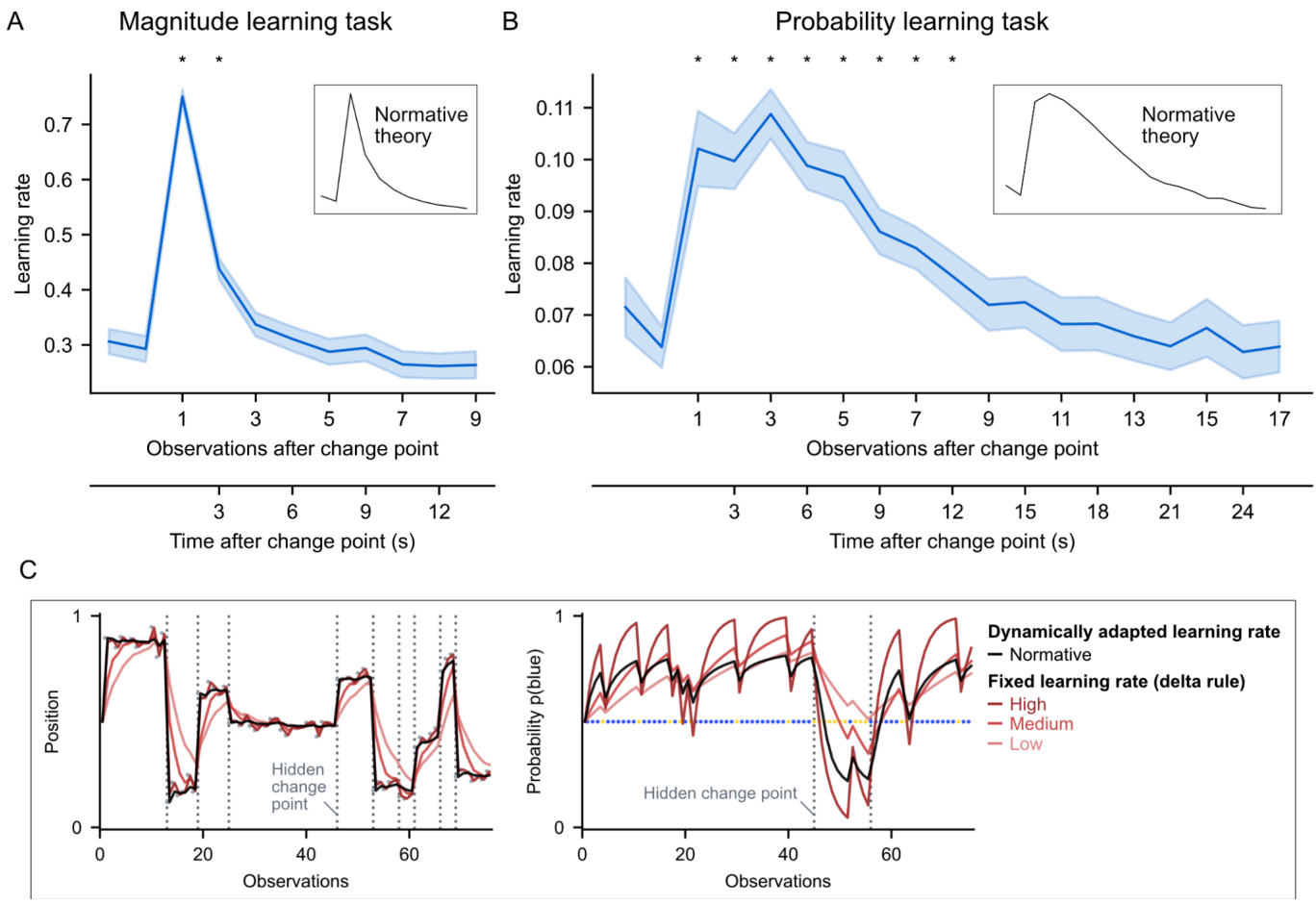


Fig. 2. Dynamics of subjects' learning rate in response to a change point in the two tasks. The subjects' learning rates were measured at each observation (equation [1]), then aligned to change points and averaged within-subject. Lines and error bands show mean ± s.e.m. across subjects. Subjects dynamically adapt their learning rate after a change point, increasing it transiently and mainly at the first observation in the magnitude learning task (A), and more prolongedly in the probability learning task (B). Stars show statistically significant differences compared to the baseline measured at the two observations before the change point ($p<0.05$, two-tailed, FWE cluster-corrected for multiple comparisons across time). The subjects' learning rate dynamics are very similar to those prescribed by normative theory, shown in the insets. (C) Examples illustrating the difference in behavior between a fixed learning rate (delta-rule, red curves, for different values of the learning rate) and a dynamic one (normative learner, black curve). When the learning rate is fixed, regardless of its value, the behavior shows systematic deviations making learning less efficient: after a change point, the estimates are too slowly updated, and in the absence of change points, they fluctuate too much following the random variability in observations. By dynamically increasing and then decreasing the learning rate after a change point, the normative learner quickly updates and then stabilizes its estimate.

Two normative determinants of the learning rate. Having demonstrated the dynamic nature of subjects' learning rate, we sought to identify the determinants that drive these dynamics (question 2). To this end, we formalized the two tasks within the same probabilistic framework and resorted to the general principles that govern learning according to normative theory. Below, we summarize the essential elements for understanding the normative determinants identified.

According to normative theory, optimal estimates are obtained by calculating the posterior distribution $p(h_t \mid x_{1:t})$, that is, the posterior distribution for the value of the hidden quantity (h_t), given the observations received so far in the sequence ($x_{1:t}$). The posterior must be updated each time a new observation is received. The normative updating process, whereby the new posterior is calculated from the previous one and the last observation, can be broken down into two steps (see Fig. 3A and equation [4] in Methods). In the first step, the previous posterior (that is, the posterior on h_{t-1}) is transformed into a prior on h_t by incorporating the generative probability of a change point occurring between the two. In the second step, this prior is multiplied with the likelihood function of the last observation to give the new posterior (that is, the posterior on h_t).

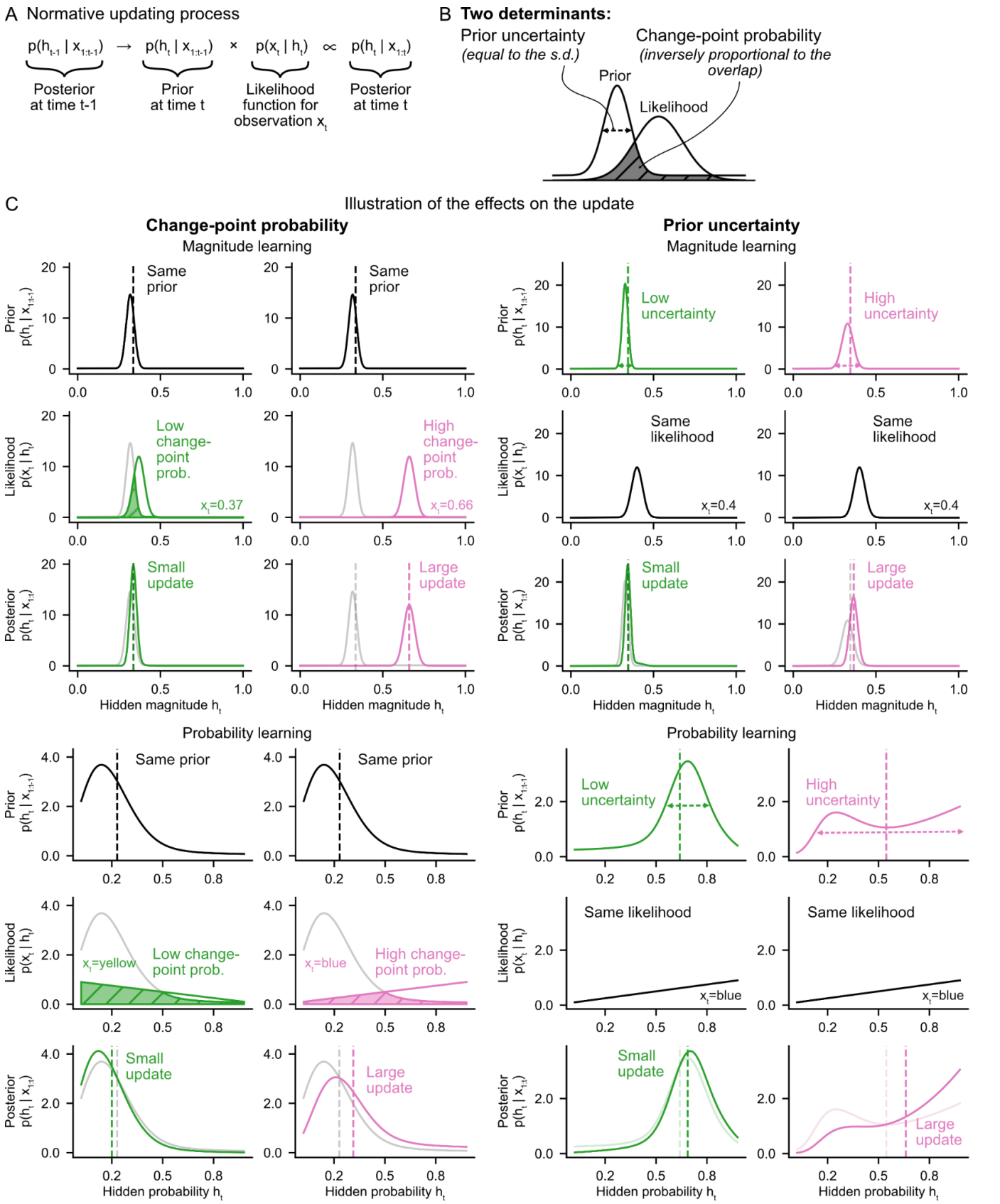


Fig. 3. Two determinants regulate updating in normative theory. (A) The normative updating process involves the product of the prior, $p(h_t|x_{1:t-1})$, and the likelihood of the last observation, $p(x_t|h_t)$. (B) The amount of update resulting from the product is determined by two factors: the uncertainty of the prior (dispersion), and the change-point probability indicated by the last observation (visualized by the degree to which the likelihood overlaps with the prior: the less it overlaps, the greater the change-point probability). (C)

Illustration of the effect on the update of each determinant in each task. The specific effect of each determinant is shown by manipulating either the likelihood (through the last observation) or the prior (through the observations prior to last). A higher change-point probability or prior uncertainty each produce a greater update (visible by the shift of the distribution and its mean between the prior and posterior). Vertical dashed lines on the plots of the distributions indicate their means. On the likelihood and posterior plots, the prior is shown in semi-transparency to visualize the overlap and the update, respectively.

Two factors determine the amount of update (and therefore, the learning rate, which is proportional to the amount of update, equation [1]) that results from the product of the prior and the likelihood function (Fig. 3B).

(i) The *prior uncertainty*, u_t, quantified by the standard deviation of the previous posterior:

$$u_t = SD[h_{t-1}|x_{1:t-1}] \qquad [2]$$

This is the uncertainty about the value of the hidden quantity before receiving the last observation. The greater the uncertainty, the larger the resulting update for the same observation, as the wider spread of the distribution leads to a larger shift (towards values where the likelihood is greater) when it is multiplied with the likelihood function. See Fig. 3C, right.

(ii) The *change-point probability*, Ω_t, which is the term used in previous literature for the estimated probability that a change point occurred at the last observation (McGuire et al., 2014; Nassar et al., 2010, 2012; Vaghi et al., 2017). It is expressed as:

$$\Omega_t = p(h_t \neq h_{t-1}|x_{1:t}) \qquad [3]$$

The change-point probability can be understood as the degree to which the likelihood function deviates from the prior. Visually, it corresponds to the degree to which the likelihood function and the prior overlap (Fig. 3B). The change-point probability is inversely proportional to this overlap (see denominator of equation [5] in Methods: the integral is equivalent to the area under the curve of the product of the two curves, which is strongly related to their area of overlap). It is also strongly related to the magnitude of the error elicited by the last observation, $x_t - v_{t-1}$, also known as prediction error (Spearman correlation $\rho>0.97$ in each of the two tasks). A higher change-point probability leads to a larger update because, as the change-point probability increases, the previous posterior is increasingly discarded (by multiplication with the likelihood function). See Fig. 3C, left.

Differential effect of the two determinants in magnitude and probability learning. Normative theory predicts a differential effect of change-point probability and prior uncertainty on the learning rate in the two tasks: the effect of change-point probability should dominate over that of prior uncertainty in the magnitude learning task, and the effect of prior uncertainty should dominate over that of change-point probability in the probability learning task (Fig. 4A). The effect sizes are measured by the weights of a multiple linear regression on the learning rate, and made comparable through z-scoring prior to regression. This differential effect is due to the fact that the variations of the two factors are not equal in the two tasks: change-point probability has much larger variations in magnitude learning compared to the probability learning, and conversely, prior uncertainty has much larger variations in probability learning compared to the magnitude learning (Fig. 4B). Thus, while both factors modulate the learning rate in both tasks, the factor that varies the most in the task is most responsible for the variations in the learning rate. The differences in variations of the two factors are related to differences in the likelihood function involved in magnitude vs. probability learning, which, in the case of magnitude, is more precise (see Fig. 3C, top vs. bottom), provides more information (Fig. S1), and thus significantly reduces uncertainty and can produce extreme values of change-point probability.

We further characterized the effects of change-point probability and prior uncertainty by relating them jointly to the learning rate and plotting the qualitative patterns that should be found in the two tasks (Fig. 4 C and D). These effects are best understood when viewed over time, by relating the variables to each other and to the observations, see Fig. 4 E and F.

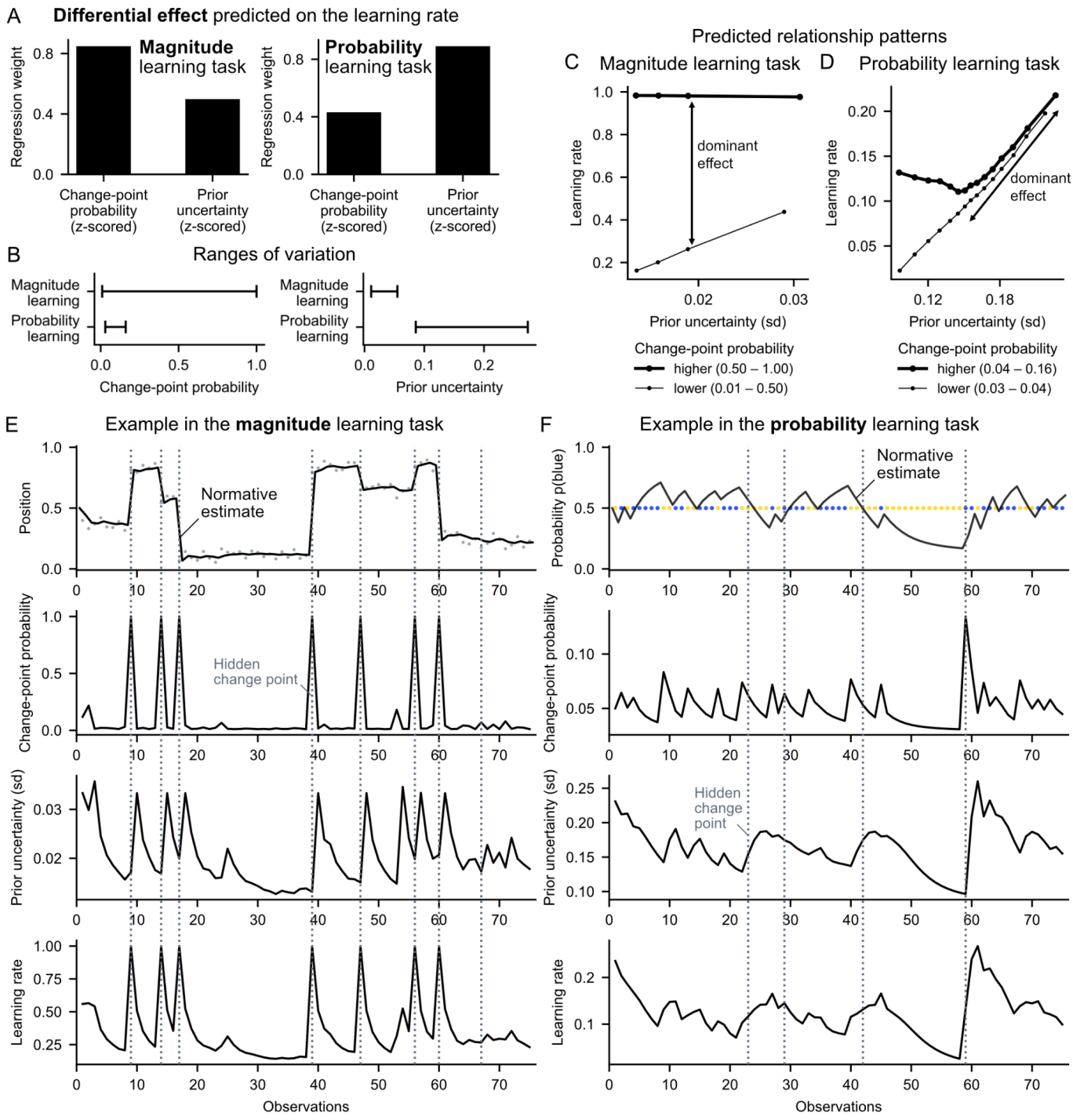


Fig. 4. Differential effect predicted by normative theory in the two tasks. (A) In the magnitude task, the effect of change-point probability on the learning rate should dominate over that of prior uncertainty, whereas in the probability task, the effect of prior uncertainty should dominate. Effects are quantified by the weights of a multiple regression of the two factors on the normative learning rate (to make the weights commensurable, all variables were z-scored). (B) Ranges of variation (min, max) of the two determinants in the two tasks. (C and D) Relationship patterns predicted between the learning rate, prior uncertainty and change-point probability in the magnitude (C) and probability (D) task. In A, C and D, we carried out the analyses on the model in the same way as for subjects in Fig. 5. (E and F) Examples illustrating the dynamics of the different normative variables as a function of the sequence in the magnitude (E) and probability (F) task; note that the y-axis greatly differs between tasks for change-point probability, prior uncertainty and learning rate.

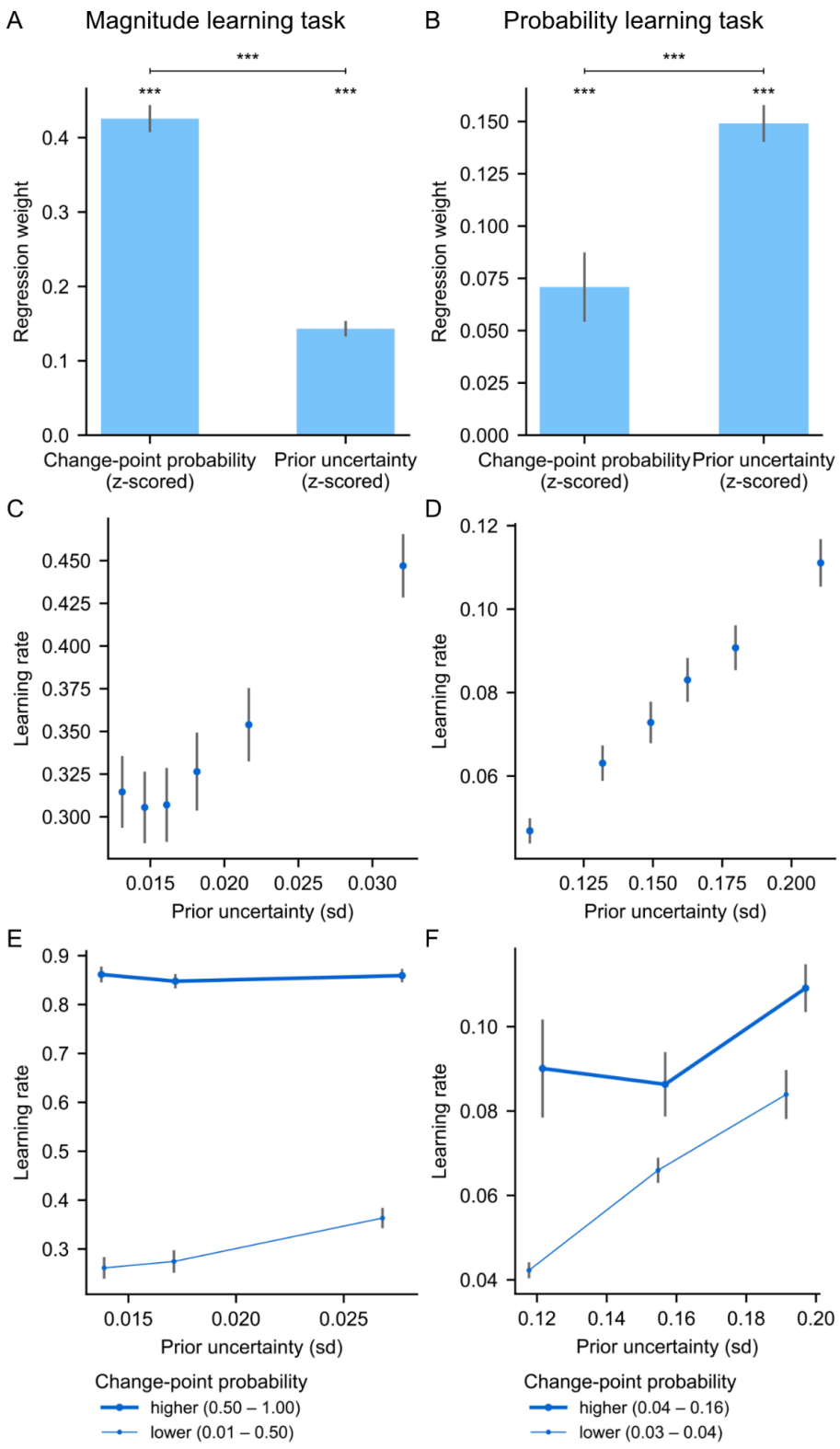


Fig. 5. Subjects' learning rates are differentially affected by the two determinants, as predicted by normative theory. (A and B) Weights of the change-point probability and prior uncertainty on subjects' learning rates in the magnitude (A) and probability (B) task, calculated per subject by multiple regression on the subject's learning rate. As predicted in Fig. 4A, the effect of change-point probability dominates in the magnitude task; that of uncertainty dominates in the probability task. ***: $p<0.001$, two-tailed t-tests. (C and D) Subjects' learning rate increases with prior uncertainty, especially in the probability task. (E and F) The interaction effects between prior uncertainty and change-point probability on subjects' learning rate are similar to the normative model (Fig. 4 C and D). In all plots, subjects' learning rates were measured at each observation (equation [1]), and prior uncertainty and change-point probability were calculated using the normative model for the same observations. Bars/dots and error bars show mean ± s.e.m. across subjects.

Do human subjects regulate their learning rate according to the two determinant factors as predicted by normative theory? To find out, we performed the same analyses as we had done with the normative model but using the subjects' learning rates instead. All the normative signatures we had identified were found in the subjects. First, the weights of change-probability and prior uncertainty on subjects' learning rates are significant in both tasks (Fig. 5 A and B) (in the magnitude task, mean±s.e.m weight and two-tailed t-test against zero for change-point probability: 0.43±0.02, t_{95}=23.4,

$p=10^{-41}$, and for prior uncertainty: 0.14±0.01, $t_{95}=13.8$, $p=10^{-24}$; in the probability task, for change-point-probability: 0.07±0.02, $t_{95}=4.3$, $p=10^{-5}$, and for prior uncertainty: 0.15±0.01, $t_{95}=17.0$, $p=10^{-30}$). Second, in the magnitude task, the weight of change-point probability is significantly higher than that of prior uncertainty (Fig. 5A; two-tailed paired samples t-test: $t_{95}=17.1$, $p=10^{-30}$; see also results in Supplementary Text 2 for comparison with previous studies). Conversely, in the probability task, the weight of prior uncertainty is significantly higher than that of change-point probability (Fig. 5B; $t_{95}=-4.4$, $p=10^{-5}$). Third, a graded co-variation of the learning rate with prior uncertainty can be observed in the magnitude task (Fig. 5C) and even more so in the probability task (Fig. 5D). Fourth, the interaction effects between prior uncertainty and change-point probability on the learning rate in subjects are similar to the normative ones (Fig. 5 E and F).

Average human learning is close to normative. So far, we have investigated the many ways in which human learning behaves normatively, but it might also have systematic deviations. Here, we studied systematicity at the group level by quantifying a systematic deviation as the degree to which the average estimate of human participants for a given observation deviates from normative estimate. Systematic deviations should be distinguished from non-systematic deviations, arising from mere variability. Both contribute to the deviations observed in subjects, but what proportion of those deviations can be attributed to the systematic vs. non-systematic component at the group level?

We disentangled the two components by performing a decomposition of the mean squared error. According to statistical theory, the mean squared error (*mse*) is the sum of two sources of error: the (squared) *bias error* (*sbe*), which is the systematic component, and the *variance error* (var), which is the non-systematic component (Hastie et al., 2009).

$$mse = sbe + var$$

Here, we are interested in the error between the normative estimate and the subjects' estimate for a given observation of the same sequence. To measure the bias and variance errors empirically in our study, we presented identical observation sequences across subjects (see Fig. 6, examples at the bottom). This allowed us to empirically calculate the mean and variance of the estimate across subjects, and the resulting bias and variance errors, for each observation of each sequence (see Methods for the mathematical expressions of this calculation). We then summed the errors across observations and sequences to obtain the proportion of the total mean squared error due to bias vs. variance. (Note that the variance measured in this way at the group level encompasses both inter-subject variability, which accounts for individual biases non-reproducible across individuals, and intra-subject variability as in (Drugowitsch et al., 2016), which is reflected across subjects in our case as there is at most one presentation per subject.)

Using this method, we found that only a small proportion of the deviations were systematic: in total across sequences, the bias/variance ratio was 23.5/76.5% (±1.1% s.e.) in the magnitude learning task and 17.3/82.7% (±0.7% s.e.) in the probability learning task (Fig. 6). This indicates that there are relatively few systematic deviations in average human learning.

We observe that there is variability in subjects' estimates, and this variability seems to be autocorrelated over time (as seen in the examples in Fig. 6). Such a variability can be produced in a model by introducing noise in the learning process. In other studies, introducing learning noise helped better explain subjects' choices, and resulted in an adjustment of choice stochasticity (also known as exploration-exploitation tradeoff) to the surprise level (Drugowitsch et al., 2016; Findling et al., 2019). These observations led us to wonder whether such a learning noise could give rise to the adjustments of the learning rate we observed in subjects. We examined this possibility through simulations of learning noise (Fig. S5), with a noise level that could be constant, as in (Drugowitsch et al., 2016), or scaled to the magnitude of the prediction error, as in (Findling et al., 2019). In both cases, these simulations did not reproduce any of the learning rate adjustments observed in subjects (Fig. S5C). Therefore, the results reported in our study regarding the subjects' ability to dynamically adapt their learning rate are not explained by learning noise.

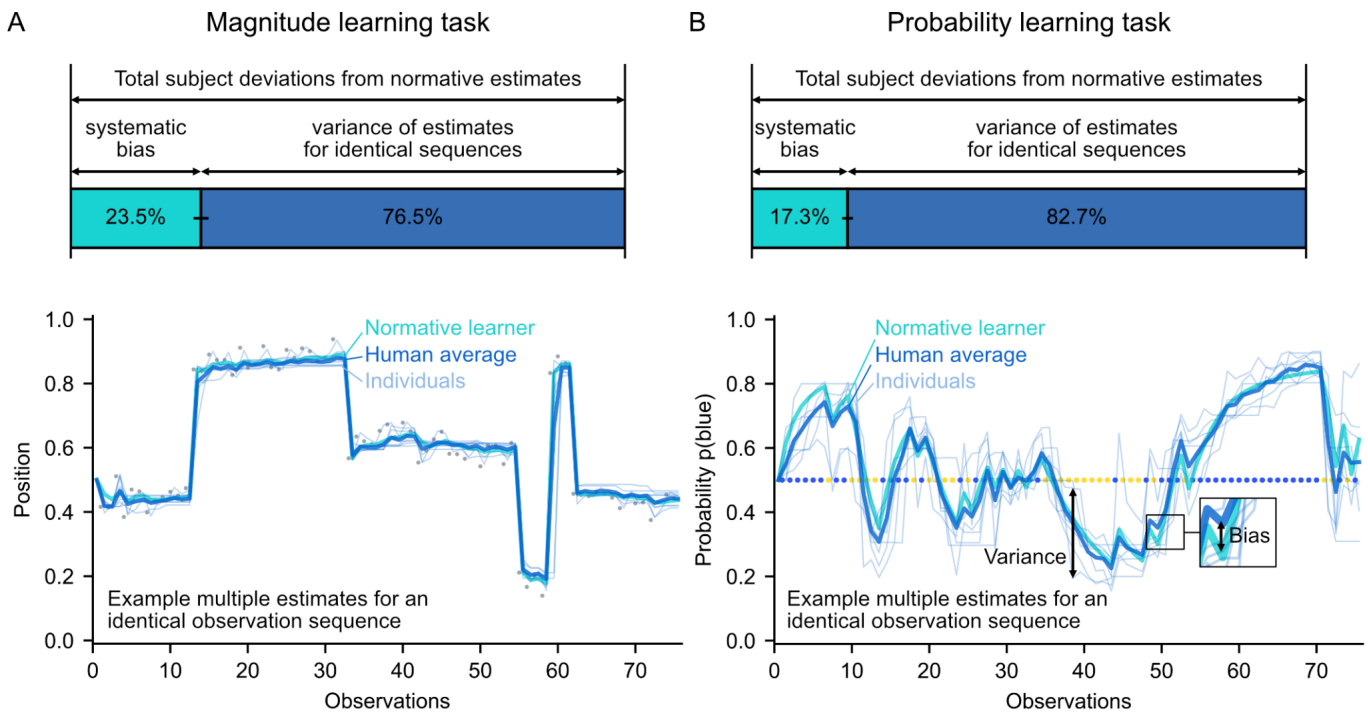


Fig. 6. Group-level decomposition of deviations from normative learning reveals minor systematic deviations compared with the overall variance. Identical observation sequences were presented across subjects. The bottom plots show, for one of the sequences in each task and each observation in the sequence, the estimates of different subjects, their average and the normative estimate. The total deviation (mean squared error) between the subjects' estimates for a given observation and the normative estimate is equal to the sum of two terms: the bias and the variance, which respectively measure systematic and non-systematic deviations across subjects. We calculated the proportion of bias/variance across sequences. The bar graphs at the top show the total proportion across sequences ± s.e. (obtained by bootstrapping) for the magnitude (A) and probability (B) task.

Discussion

Many studies in psychology and neuroscience use a fixed learning rate to model the subject's behavior (with models known by various names such as delta rule, Rescorla-Wagner, temporal difference (TD) learning, Q-learning, SARSA...) (O'Doherty et al., 2003; Rescorla & Wagner, 1972; Rosenblatt, 1961; Schultz et al., 1997; Sutton & Barto, 2018). Our results actually show that in humans, the learning rate varies dynamically from one observation to the next based on what is observed, and is normatively adapted in response to inferred changes in the environment (Fig. 2). We found evidence for these normative dynamic adjustments of the learning rate in two common types of learning, namely magnitude learning and probability learning, despite fundamental differences that exist between these two types of learning, demonstrating the generality of dynamic adaptive learning in humans.

The dynamic adaptation we have shown for probability learning is distinct from an average adaptation to global environmental statistics such as volatility. In the latter, the average learning rate was found to be higher in a block of trials containing many change points (volatile block) than in a block containing no change points (stable block) (Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019). Here, the adjustment of the learning rate we report is not on average across trials but from trial to trial, in response to single change points. This dynamic adaptation implies an adaptation to volatility (because more volatile environments have more change points, inducing more frequent increases in the learning rate and thus a higher average learning rate), but the reverse is not necessarily true. For instance, adaptation to volatility can arise from a switch between strategies that are tailored to different volatility conditions, with each strategy having its own static learning rate; for example, in a choice task, a strategy closer to choice repetition can be used in the stable condition and one closer to win-stay, lose-shift in the volatile condition (Cook et al., 2019).

The generality of our finding is further strengthened by several advantages of our paradigm compared to the numerous probability learning paradigms based on choices used previously, such as probabilistic reversal learning paradigms

(Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019; Cools et al., 2002). First, our measure of the subjects' learning rate does not require any model. It is obtained directly from the subjects' behavior (equation [1]). It is therefore not sensitive to the researcher's modeling assumptions, as is the case when the learning rate is estimated by fitting a model parameter across trials. Second, change points in our tasks are not limited to simple reversals and the times at which they occur are completely unpredictable). This eliminates the possibility that subjects may have anticipated the change points (as do some models of reversal tasks (Costa et al., 2015; Wilson et al., 2014)).

Before further discussing the implications of our findings, we would like to highlight the context in which they were obtained. Here, we aimed to study how subjects learn the hidden variable (a magnitude or a probability) of a given generative process, not how they learn the structure of a generative process. Therefore, we provided subjects with detailed instructions so that they fully understood the generative process of the task, and the goal of estimating a time-varying variable. Learning in a context where the structure is unknown, or partially known (to the extent of the instructions), is a different problem and subject of study known as structure learning (Gershman & Niv, 2010). It requires a distinct theoretical treatment and needs its own empirical investigation. The extent to which our findings apply to structure learning remains to be determined.

Given that dynamic adaptive learning has been demonstrated in our study, research in neuroscience should now uncover its neural mechanisms. To study this question, our paradigm could be reused in neuroimaging experiments, and as a starting point, neural measurements could be related to the computational determinants we presented. Catecholaminergic neuromodulators (noradrenaline and dopamine) are good candidates for investigation as several theoretical and empirical works indicate that they may play a role in regulating learning based on surprise and uncertainty (Aston-Jones & Cohen, 2005; Yu & Dayan, 2005). These neuromodulators provide a biological implementation of gating, also known as gain modulation, a computational mechanism that we have previously shown to play a crucial role for artificial neural networks to achieve the very same dynamic adaptations that we have found here in humans (Foucault & Meyniel, 2021).

Differences between magnitude and probability learning. In light of the differential effects we found in the magnitude and probability learning tasks (Fig. 5), it will be important in future studies of learning to distinguish the contribution of, on the one hand, the effects related to the immediately received observation and the quantities associated with it, such as error magnitude, surprise and change-point probability, and on the other hand, the effects of prior uncertainty arising from past observations. Moreover, it will be important to distinguish the role of these two types of effects in magnitude learning vs. probability learning, as they do not necessarily involve the same mechanisms in these two types of learning. In the present study, we found that there was overall little correlation between the magnitude and probability task across subjects in terms of how much they weighted change-point probability (no correlation, r=-0.01) and prior uncertainty (significant but weak correlation, r=0.26), compared to e.g. how frequently they updated their reports (r=0.63) (see Supplementary Text 4 for more details). This invites us to be particularly careful and to not consider a priori that the results discovered in one type of learning will generalize to all types of learning under uncertainty, because it runs the risk of confusing what are potentially very different processes.

Comparing previous fMRI studies conducted separately on magnitude learning (McGuire et al., 2014) and probability learning (Bounmy et al., 2023; Meyniel & Dehaene, 2017), we noted differences between the two in the neural correlates that were specific to each type of effect. Specifically, when examining the selective effects of the error (quantified by change-point probability or surprise), we noted that frontal regions were implicated mainly in probability learning (right superior frontal gyrus, left and right supplementary and cingulate eye fields). In magnitude learning, they involved primarily medial regions of the visual cortex (calcarine sulci), whereas in probability learning the effects in visual cortex were more lateral (left/right V3/V4). When examining the selective effects of prior uncertainty (quantified by relative uncertainty or confidence), we noted that ventromedial and anterior prefrontal cortex regions were involved primarily in magnitude learning. Interestingly, both types of learning showed neural correlates in the parietal cortex, but in magnitude learning these correlates were specific to prior uncertainty, whereas in probability learning they were common to both prior uncertainty and surprise.

In general, several factors may give rise to differences between the two types of learning. First, there may be distinct neural mechanisms for processing different types of quantities (magnitudes vs. probabilities). Second, the ranges and dynamics of change-point probability and prior uncertainty are different in magnitude and probability learning (Fig. 4), and different neural substrates may be recruited in these different regimes. Third, these two factors may be processed

differently in the two types of learning because, in magnitude learning, change-point probability is more diagnostic than prior uncertainty for detecting change points, while in probability learning, prior uncertainty is a more reliable signal for inferring a change point in the recent past (Fig. 4 E and F). In addition, there may be a trade-off in the degree of processing one can allocate to compute and use these two factors that favors the most relevant factor in a given learning context, namely the change-point probability when learning magnitudes and the prior uncertainty when learning probabilities (Fig. 4A, 5 A and B).

Frequent updating during the learning process. A previous study proposed that the brain does not update its probability estimate at each observation, but only on rare occasions (Gallistel et al., 2014). This conclusion was based on an empirical observation: in Gallistel et al.'s study, subjects overtly updated their reported probability estimate only once every 18.1 observations on average, with only 7% of updates occurring after a single observation. However, a lack of updating in the subject's report may not necessarily reflect a lack of updating in the subject's internal learning process. In contrast to Gallistel et al.'s study, our study shows frequent updating during the learning process. Subjects updated their estimate at each observation in the vast majority of cases: 84% of updates occurred after a single observation (vs. 7% in Gallistel et al.), and on average subjects updated their estimate every 1.4 observation (vs. 18.1) (Fig. S3). This makes our data on human probability learning rather novel.

This difference in the frequency of report updates between Gallistel et al.'s study and our study can be explained by task differences: the cost for the subject to overtly update their estimate greatly varies depending on how the task is designed, and when this cost is high, subjects may not manifest their update. In Gallistel et al.'s task, the action required to update the report induces a time cost which, by reproducing the task interface, we estimated to be 4 to 5 s per observation. If subjects had performed this action for all 1,000 observations of a session, it would have taken them more than an hour (estimated 67 to 83 min) per session, instead of the 25.6 min they took on average (Gallistel et al., 2014). In our task, there is no time cost since the observations occur at fixed time intervals. This design choice seems to have made subjects much more willing to express their update. Subjects in our study also received a performance bonus encouraging them to report their updates.

The frequency of subjects' updating is critical to the results we have shown on the dynamics of the learning rate at the scale of an observation. Note that all our conclusions hold after stringently excluding from analysis all trials in which the report was not updated (Fig. S4).

Computational models of human learning. Although the human learning algorithm remains to be determined, the results shown here impose strong computational constraints on this algorithm. Models of this algorithm should be able to exhibit the same dynamic adaptations as humans do in response to change points and depending on change-point probability and prior uncertainty (Fig. 2 and 5). They should also be able to reproduce the variability of human learning, and the fact that it is on average close to normative (Fig. 6).

Various models have been proposed in the literature, including the Kalman filter (Kalman, 1960), the Hierarchical Gaussian Filter (HGF) (Mathys et al., 2011), a model estimating the volatility and stochasticity of the environment (assumed to follow a random walk as in the Kalman filter) (Piray & Daw, 2021), the proportional-integral-derivative (PID) controller (Ritz et al., 2018), an adaptive mixture of delta-rules (Wilson et al., 2013), a model with metaplastic synapses guided by a change detection system (Iigaya, 2016), the normative model we used here (which computes the optimal solution via Bayesian inference) (Adams & MacKay, 2007; Behrens et al., 2007; Heilbron & Meyniel, 2019), particle filters (which uses sampling to approximate the optimal solution) (Brown & Steyvers, 2009; Prat-Carrabin et al., 2021), and several architectures of recurrent neural networks (RNNs) (Foucault & Meyniel, 2021; Wang et al., 2018). The Kalman filter is a particular case because it qualifies for level 1 but not level 2 of adaptive learning as we presented in the introduction: its learning rate depends on the global level of stochasticity and volatility in the environment, but does not depend on what is observed (the changes in learning rate over time are simply dictated by the number of observations). The other models qualify for level 2 (they dynamically adapt their learning rate), but it remains to be evaluated to what extent they can, in both types of learning, conform to the specific normative properties demonstrated here in humans.

In a previous article (Foucault & Meyniel, 2021), we have shown that a small recurrent neural network (as small as three recurrent units) could solve the probability learning task quasi-optimally and reproduce all the above-mentioned normative properties (Fig. 2 to 5). This model thus provides a low-cost, neurally feasible solution. In addition, it has the

advantage of not incorporating a generative model of the environment a priori: through mere exposure to the environment, the network spontaneously develops its dynamic adaptive learning capabilities (which is sometimes referred to as 'meta-learning') (Foucault & Meyniel, 2021; Wang et al., 2018). One mechanistic insight from our study was that small networks demonstrated optimal adaptive learning capabilities only when equipped with a gating mechanism. This is interesting to put in perspective with the other proposed models as they often involve multiplicative interactions (i.e. multiplication between two state variables or a state variable and an input variable) like gating.

Using our behavioral dataset, future studies will be able to evaluate and compare models to probe the algorithms underlying human learning. New data can also be collected easily and in large amounts using the experimental paradigm that we make available to the community (1 report per observation every 1.5 s).

Relevance for artificial intelligence. Using a dynamic adaptive learning rate is relevant for machine learning and artificial intelligence. The two determinants we identified could guide the search for new efficient, adaptive learning methods. One advantage of change-point probability and prior uncertainty, compared to state of the art methods such as Adam, AdaGrad and RMSprop (Goodfellow et al., 2016), is that they do not require to evaluate the cost function of the task, which in many cases is not accessible at the level of one input sample (in our task for example, the cost function is the error between the subject's estimate and the true quantity, which is hidden). This is especially relevant for learning from few input samples, as in one-shot or few-shot learning. According to our findings, change-point probability might be particularly relevant for one-shot learning in contexts where individual samples are highly informative, and prior uncertainty for few-shot learning in contexts where individual samples are less informative. This could be linked to current research in machine learning aimed at improving systems' ability to assess their uncertainty (Gal & Ghahramani, 2016; Hüllermeier & Waegeman, 2021; Kendall & Gal, 2017; Kompa et al., 2021).

Methods

Participants.

96 human subjects participated in our study (38 female, median age 30 years, interquartile range 25–37 years, from 19 different nationalities). Participants were recruited from Prolific, an online platform for recruiting participants focusing on academic research. The study was approved by the Comité d'Ethique de la Recherche of the Paris Saclay university (#CER-Paris-Saclay-2023-010). Participants were required to use a computer with a touchpad or mouse and a screen large enough to perform the task. They were rewarded £8.40 for their participation (£9/h at an a priori estimated completion time of 56 min, which also turned out to be the median completion time of the participants), plus a performance-based bonus of up to £4.20 (50% of the base pay). Informed consent was collected before the start of the experiment. No participants were excluded from the analyses.

Ethics Statement.

The study has been approved by the Ethics Committee of the Paris-Saclay University (Comité d'Ethique de la Recherche, approval #CER-Paris-Saclay-2023-010). Participants gave their written informed consent prior to participating in the study.

Experiment steps.

The experiment consisted of the following steps: Task 1 instructions, Task 1 sessions, Task 2 instructions, Task 2 sessions. Each task session consisted of a sequence of 75 observations, taking under 2 min to perform (see examples Fig. 1 E and F). In total, the subjects took a median time of 56 min to complete the whole experiment, including instructions (interquartile range 52–65 min). They performed twenty-one task sessions (six of the magnitude task and fifteen of the probability task; this split was determined a priori with an independent pilot study of eight subjects in order to obtain a sufficiently reliable measure of a subject's learning rate adjustments, with approximately equal relative error in the measurement in both tasks). At the end of each task session, subjects received feedback whose purpose was to maximize subjects' engagement with the task. This feedback did not seem to induce a training effect, as performance was stable over the course of the task (Fig. S6). The feedback display showed the subject's score for the session and the corresponding monetary gain (proportional to the score), and also showed the true values of magnitude/probability along with the subject's estimates. The score was calculated based on the mean absolute error between subjects' estimates and the true values of magnitude/probability. The function mapping the error to the score was designed to:

normalize the score in 0–100%; keep the score of the normative learner at a constant level (close to 100%); prevent too low scores (equal or close to 0%) that would demotivate subjects, using a softplus nonlinearity.

Task design.

The graphic elements of the tasks were designed with several goals in mind: a) to allow subjects to easily monitor the stimuli as they are appearing and adjust their estimate at the same time, b) to strike a good balance between speed and accuracy of adjustments, c) to share as much as possible between the two tasks (see Fig. 1 A and B and links below to run the tasks). The total length of the slider track was 640 CSS pixels, such that the leftmost and rightmost reachable position of the slider was at about 6–7º of visual angle from the center (1 CSS px corresponds to 0.0213º visual angle at the typical viewing distance of the user's display, https://www.w3.org/TR/css-values-3/#reference-pixel). This allowed the subject to see the slider and the stimuli at the same time. Tick marks were displayed at every 10% length of the slider to make location easier for subjects. These elements were shared between the two tasks. Additionally, there were a few design elements specific to the probability learning task: labels were put below the tick marks showing the percentage corresponding to the estimated probability (in the magnitude task, this is irrelevant as the position itself is the estimate); the portions of the slider between 0–10% and 90–100% were hatched to indicate that the hidden probability never lies within these intervals; a small yellow wheel and blue wheel were shown at the left and right edge of the slider track, respectively, to indicate the direction of estimation in relation to the two colors; we chose the blue and yellow colors to be easily recognizable by humans even with color-blindness.

Within the 1.5 s interval separating the onset of each observation, the stimulus offset occurred at 1.3s, and in the 1.3–1.5 s inter-stimulus interval, the slider thumb was highlighted (lighter stroke color) to indicate to subjects when their estimate was recorded for calculating their performance.

Throughout the session, subjects could at any time move the slider via motion tracking to update their estimate. Motion tracking was implemented by continuously tracking movements of the subject's pointer (corresponding to finger movements on a touchpad, or to mouse movements). The pointer itself was hidden during the task, to make it clear to the subject that they were controlling the slider.

In the task instructions, for the magnitude learning task, the choice of the cover story (a hidden person throwing snowballs) was inspired by (Prat-Carrabin et al., 2021).

The tasks can be freely tested online, separately at https://run.pavlovia.org/cedricfoucault/ada-learn-task/ada-pos-study.html?skipConsent (magnitude learning task) and https://run.pavlovia.org/cedricfoucault/ada-learn-task/ada-prob-study.html?skipConsent (probability learning task), and the combined experiment can be tested at https://run.pavlovia.org/cedricfoucault/ada-learn-task/ada-pos-prob-study.html (magnitude task first) and https://run.pavlovia.org/cedricfoucault/ada-learn-task/ada-prob-pos-study.html (probability task first).

observation sequences and sequence-generating processes of the tasks.

We generated a set of 100 sequences for the magnitude learning task and 150 sequences for the probability learning task. Each time a subject performed the tasks, they were shown sequences randomly sampled without replacement from these sets.

For the sake of clarity and to be able to compare the two tasks, we use here as the reference unit the horizontal position within the slider interval normalized between 0 and 1 (left and right edge of the slider track, respectively). This unit applies to the estimates (slider positions) and the hidden magnitudes/probabilities these estimates are about, and to the observations (for the probability task, observation 'yellow' and 'blue' correspond to '0' and '1', respectively, since when the estimate is equal to 1 this corresponds to a 100% probability of the observation being blue). For the magnitude learning task, the observation-generating process was the same as that used by (Nassar et al., 2012) in order to replicate their procedure: the observations were drawn from a Gaussian distribution with a standard deviation of 10/300 (in normalized unit), and whose mean (the hidden magnitude) changed with a probability of zero for the first three observations following a change point and 1/10 for all trials thereafter. For the probability learning task, the observations were drawn from a Bernoulli distribution whose parameter (the hidden probability) changed with a probability of zero for the first six observations following a change point and 1/20 for all trials thereafter. The probability was sampled uniformly between 0.1 and 0.9 initially, and resampled uniformly in the same interval at each change

point subject to the constraint that the resulting change in the odds p/(1-p) be no less than fourfold. These parameters were chosen to maintain continuity with previous studies (Bounmy et al., 2023; Heilbron & Meyniel, 2019; Meyniel et al., 2015; Meyniel & Dehaene, 2017). Sampling probabilities between 0.1 and 0.9 avoids producing sequences with excessively long streaks of identical observations where almost no update needs to be made. The minimum distance between change points and the odds-change constraint avoid having change points that are too short-lived or too subtle to be perceived (even the optimal solution shows almost no learning rate adjustments in response to such changes). Other studies on probability learning also used probabilities between 0.1 and 0.9 and constrained change points such that nearly imperceptible change points did not occur (for comparison, the often-used reversal between 20% and 80% produces an odds change of 16) (Behrens et al., 2007; Browning et al., 2015; Cook et al., 2019).

Note that, for the magnitude learning task, we used the same parameter values as previous studies (Nassar et al., 2012; Vaghi et al., 2017). Equalizing the two tasks in terms of the informative value of the observations would not be practically feasible (see also Fig. S1).

Behavioral analyses.

The subject's estimate v_t, for observation x_t, was taken as the normalized position of the slider at the end of the 1.5s interval just before receiving the next observation (the estimate prior to the first observation, v_0, is 0.5, as the slider is initially positioned in the middle). The value x_t is equal to, for the magnitude task, the (normalized) position of the stimulus, and for the probability task, 1 for a blue draw and 0 for a yellow draw. v_{t-1} and v_t are the subject's estimate prior to and after receiving that observation, and α_t is the subject's learning rate for that observation.

In the results relating to the subjects' estimate accuracy, the significance was calculated by computing the Pearson correlation at the subject-level between the subject's and the normative estimates, and then performing at the group-level a two-tailed, one-sample t-test of those correlations against zero.

When analyzing learning rates in the magnitude learning task, similar to (Nassar et al., 2012; Vaghi et al., 2017), we ignored from the analyses outliers in the learning rates (which might occasionally occur when the error is very close to 0), corresponding to learning rates above 1.3 or below -0.6. These values were determined a priori using the distribution of the subjects' learning rates in an independent pilot study of eight subjects by taking the values above and below which the empirical density was less than 0.05.

Normative models.

Normative models are Bayesian models that compute the posterior distribution $p(h_t \mid x_{1:t})$, which is the probability distribution over the latent quantity that generates the observations (h_t) given the observations received by the subject ($x_{1:t}$), with knowledge of the true generative process of observations of the task (Ma et al., 2023).

Given the properties of the generative process in both tasks, this distribution can be computed sequentially using the following update equation (this method is known as Bayesian filtering, for a textbook, see (Särkkä & Svensson, 2023)).

$$p(h_t \mid x_{1:t}) \propto p(x_t \mid h_t)\, p(h_t \mid x_{1:t-1})$$

$$\propto p(x_t \mid h_t) \int p(h_t \mid h_{t-1})\, p(h_{t-1} \mid x_{1:t-1})\, dh_{t-1} \quad [4]$$

where $\propto$ denotes equality up to a normalization constant (the left-hand side is obtained from the right-hand side by dividing the right-hand side by its sum over the possible values of h_{t+1}, so that the distribution sums to 1). This equation is derived by applying the rules of probability theory and leveraging two conditional independence properties of the generative process: h_{t+1} is conditionally independent of $x_{1:t}$ given h_t, and x_{t+1} is conditionally independent of $x_{1:t}$ given h_{t+1}.

Having computed the posterior distribution, its mean can be computed to obtain the model's estimate, $v_t = E[h_t \mid x_{1:t}]$, and its standard deviation to obtain the uncertainty, which is the prior uncertainty for the next time step ($u_t = SD[h_{t-1} \mid x_{1:t-1}]$, equation [2]). The change-point probability can be computed from the posterior distribution using the following equation, which is derived from Bayes' rule by leveraging the two conditional independence properties mentioned above:

$\Omega_t = p(h_t \neq h_{t-1} | x_{1:t})$

$= p(x_t \mid h_t \neq h_{t-1})\, p(h_t \neq h_{t-1}) \,/\, \int p(x_t \mid h_t)\, p(h_t \mid x_{1:t-1})\, dh_t$ [5]

Below, we provide the details of the specific implementation of the model we used for each task.

Probability learning task. We reused the implementation of the Bayesian model for the probability learning task of (Meyniel, 2020) whose code is available online at https://github.com/florentmeyniel/TransitionProbModel. We made one modification to match the task asked of the subject in the present study: as the subject was told that the hidden probabilities are drawn in [0.1, 0.9] (with the restriction of the slider to that interval), we also incorporated this knowledge into the model, by setting the corresponding distribution, $p(h_1)$ and $p(h_{t+1} \mid h_{t+1} \neq h_t)$, equal to the uniform distribution on [0.1, 0.9], rather than [0, 1] as in the original version.

Magnitude learning task. We implemented the reduced Bayesian model of the magnitude learning task described in (Nassar et al., 2012), which has also been used in other magnitude learning task studies (McGuire et al., 2014; Vaghi et al., 2017). We implemented this reduced model rather than the full model to replicate the procedure of Nassar et al, 2012, whose magnitude learning task serves as a reference for comparison with the probability learning task. This model simplifies computations by only maintaining the first two moments of the posterior distribution, which was reported to have minimal effect on estimates in this task (Nassar et al., 2010, 2012). The model variables are computed sequentially using the following system of equations:

The estimate of the latent mean:

$b_{t+1} = b_t + \eta_t\,(x_t - b_t)$

$v_t = b_{t+1}$

Learning rate:

$\eta_t = \tau_t + (1-\tau_t)\,\Omega_t$

The change-point probability:

$\Omega_t = \mathcal{U}(x_t)\, H \,/\, [\mathcal{U}(x_t)\, H + \mathcal{N}(x_t;\, b_t,\, \sigma_t^2)\,(1-H)]$

The predictive variance of observations:

$\sigma_t^2 = N^2 + \tau_t N^2 \,/\, (1 - \tau_t)$

The relative uncertainty about the latent mean with respect to the predictive variance of observations:

$\tau_{t+1} = [N^2\,\Omega_t + (1-\Omega_t)\,\tau_t\,N^2 + \Omega_t\,(1-\Omega_t)\,(x_t\,\tau_t + b_t\,(1-\tau_t) - x_t)^2] \,/\, [N^2\,\Omega_t + (1-\Omega_t)\,\tau_t\,N^2 + \Omega_t\,(1-\Omega_t)\,(x_t\,\tau_t + b_t\,(1-\tau_t) - x_t)^2 + N^2]$

The uncertainty about the latent mean:

$u_t^2 = \tau_t\,\sigma_t^2$

where $\mathcal{U}$ is the uniform distribution, $\mathcal{N}$ is the normal distribution with the mean and variance given after the semicolon, N is the standard deviation of the observation generation distribution of the task generative process, and $H = p(h_t \neq h_{t-1})$ is the probability of occurrence of a change point of the task generative process.

Bias/variance decomposition of the mean squared error.

Having presented the same sequences multiple times across subjects in our study, we have, for a given sequence k out of K sequences, $N(k)$ subjects who provided estimates (K=100 and 150 sequences for the magnitude and probability tasks respectively; median $N(k)$ across sequences: 6 and 10 for the magnitude and probability tasks respectively). For a time step t of the sequence, the mean squared error (*mse*) between the normative estimate (v_t^n) and the subjects' estimate ($v_t^{s(i)}$, where $s(i)$ denotes the i-th subject among the $N(k)$ subjects) is written as:

$$mse = \frac{1}{N(k)} \sum_{i=1}^{N(k)} (v_t^n - v_t^{s(i)})^2$$

This is equal to the sum of the squared bias error (also called “bias error”) and the variance (also called “variance error”) (Hastie et al., 2009). The squared bias error (*sbe*) is written as:

$$sbe = (v_t^n - v_t^s)^2$$

where $v_t^s = \frac{1}{N(k)} \sum_{i=1}^{N(k)} v_t^{s(i)}$ is the mean estimate across subjects.

The variance (*var*) is written as:

$$var = \frac{1}{N(k)} \sum_{i=1}^{N(k)} (v_t^s - v_t^{s(i)})^2$$

We quantified the precision of our measurement of the proportion of bias/variance (and of the proportion of conservatism bias in Table S3) by the standard error (s.e., which corresponds to the margin of error of the 68% confidence interval of the measurement). To estimate it, we used a bootstrapping procedure. We generated 10,000 new sets of sequences of the same size as the original set by resampling the original set with replacement, and performed the measurement on each of those sets. The estimate of standard error is the standard deviation of the measurement across the sets.

Data and code availability

All data, and code to reproduce the results will be made freely available at the time of publication on OSF, and on our GitHub repository.

The tasks can also be readily tested online at the links given above.

Acknowledgements

We gratefully acknowledge the support received for this work. C.F. was supported by a PhD fellowship from ENS Paris-Saclay (France). F.M. was supported by the European Research Council (ERC grant #947105) and the Inserm. We thank Maëva L'Hôtellier and Alexander Paunov for useful feedback on the project. We also thank them as well as Steven Geysen and Tiffany Bounmy for piloting the experiment.

Author contributions

CF: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing—original draft preparation, Writing—review & editing; FM: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing—original draft preparation, Writing—review and editing.

Competing interests

The authors declare no competing interests.

References

Adams, R. P., & MacKay, D. J. C. (2007). *Bayesian Online Changepoint Detection*. https://arxiv.org/abs/0710.3742

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), Article 1. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221.

Bounmy, T., Eger, E., & Meyniel, F. (2023). A characterization of the neural representation of confidence during probabilistic learning. *NeuroImage*, 119849.

Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology*, *58*(1), 49–67.

Browning, M., Behrens, T. E., Jocham, G., O'reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596.

Cook, J. L., Swart, J. C., Froböse, M. I., Diaconescu, A. O., Geurts, D. E., Den Ouden, H. E., & Cools, R. (2019). Catecholaminergic modulation of meta-learning. *Elife*, *8*, e51439.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *22*(11), 4563–4567.

Costa, V. D., Tran, V. L., Turchi, J., & Averbeck, B. B. (2015). Reversal learning and dopamine: A bayesian perspective. *Journal of Neuroscience*, *35*(6), 2407–2416.

Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, *121*(3), Article 3. https://doi.org/10.1037/a0037010

Drugowitsch, J., Wyart, V., Devauchelle, A.-D., & Koechlin, E. (2016). Computational precision of mental inference as critical source of human choice suboptimality. *Neuron*, *92*(6), 1398–1411.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over-and underconfidence: The role of error in judgment processes. *Psychological Review*, *101*(3), 519.

Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., & Wyart, V. (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nature Neuroscience*, 1–12. https://doi.org/10.1038/s41593-019-0518-9

Foucault, C., & Meyniel, F. (2021). Gated recurrence enables simple and accurate sequence prediction in stochastic, changing, and structured environments. *eLife*, *10*, e71801. https://doi.org/10.7554/eLife.71801

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *International Conference on Machine Learning*, 1050–1059. http://proceedings.mlr.press/v48/gal16.html

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review*, *121*(1), 96.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, *20*(2), 251–256.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning book. *MIT Press*, *521*(7553), 800.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer.

Heilbron, M., & Meyniel, F. (2019). Confidence resets reveal hierarchical adaptive learning in humans. *PLoS Computational Biology*, *15*(4), e1006972.

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, *138*(2), 211.

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, *110*, 457–506.

Iigaya, K. (2016). Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *eLife*, *5*. https://doi.org/10.7554/eLife.18073

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), Article 1. https://doi.org/10.1115/1.3662552

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, *30*.

Kompa, B., Snoek, J., & Beam, A. L. (2021). Second opinion needed: Communicating uncertainty in medical machine learning. *NPJ Digital Medicine*, *4*(1), 4.

Lee, S., Gold, J. I., & Kable, J. W. (2020). The human as delta-rule learner. *Decision*, *7*(1), 55.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*. https://doi.org/10.1017/S0140525X1900061X

Ma, W. J., Kording, K. P., & Goldreich, D. (2023). *Bayesian Models of Perception and Action: An Introduction*. MIT press.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, 39. https://doi.org/10.3389/fnhum.2011.00039

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, *84*(4), 870–881.

Meyniel, F. (2020). Brain dynamics for confidence-weighted learning. *PLOS Computational Biology*, *16*(6), Article 6. https://doi.org/10.1371/journal.pcbi.1007935

Meyniel, F., & Dehaene, S. (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proceedings of the National Academy of Sciences*, *114*(19), E3859–E3868.

Meyniel, F., Schlunegger, D., & Dehaene, S. (2015). The sense of confidence during probabilistic learning: A normative account. *PLoS Computational Biology*, *11*(6), e1004305.

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040.

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, *30*(37), 12366–12378.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press. https://books.google.com/books?hl=en&lr=&id=sLetNgiU7ugC&oi=fnd&pg=PR5&dq=Oaksford+chater+2007&ots=IpPeO8etuI&sig=rn3Lll7WFLhUcI33o2vFqJLCI1E

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329–337.

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*(3), 346.

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications*, *12*(1), 6587.

Prat-Carrabin, A., Wilson, R. C., Cohen, J. D., & da Silveira, R. A. (2021). Human Inference in Changing Environments With Temporal Structure. *Psychological Review*, *128*(5), 879–912. https://doi.org/10.1037/rev0000276

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, *2*, 64–99.

Ritz, H., Nassar, M. R., Frank, M. J., & Shenhav, A. (2018). A Control Theoretic Model of Adaptive Learning in Dynamic Environments. *Journal of Cognitive Neuroscience*, *30*(10), Article 10. https://doi.org/10.1162/jocn_a_01289

Rosenblatt, F. (1961). *Principles of neurodynamics. Perceptrons and the theory of brain mechanisms*. Cornell Aeronautical Lab Inc Buffalo NY.

Särkkä, S., & Svensson, L. (2023). *Bayesian filtering and smoothing* (Vol. 17). Cambridge university press.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience*, 1.

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

Vaghi, M. M., Luyckx, F., Sule, A., Fineberg, N. A., Robbins, T. W., & De Martino, B. (2017). Compulsivity reveals a novel dissociation between action and confidence. *Neuron*, *96*(2), 348–354.

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., Hassabis, D., & Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, *21*(6), 860.

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2013). A mixture of delta-rules approximation to bayesian inference in change-point problems. *PLoS Computational Biology*, *9*(7), Article 7. https://doi.org/10.1371/journal.pcbi.1003150

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*, *81*(2), Article 2. https://doi.org/10.1016/j.neuron.2013.11.005

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), Article 4. https://doi.org/10.1016/j.neuron.2005.04.026

Zhu, J.-Q., Sanborn, A. N., & Chater, N. (2020). The Bayesian sampler: Generic Bayesian inference causes incoherence in human probability judgments. *Psychological Review*, *127*(5), 719.

Supplementary information

Supplementary Text and Tables

Supplementary Text 1: Comparison between normative estimates and subjects' estimates.

Linear regression.

To see how subjects' estimates compared to normative estimates, we performed a linear regression between the two at the subject level, and then summarized the results at the group level. We collected two measures derived from the regression: the Pearson correlation coefficient and the slope of the linear regression. The results are reported in Tables S1 and S2 below.

The Pearson correlation coefficient results (Table S1) had already been reported in the main text. This coefficient measures the strength of the linear relationship between normative estimates and subjects' estimates. The fact that the coefficient is significantly greater than 0 shows that subjects' estimates covary with the optimal estimates, indicating that subjects perform the task adequately.

The slope of the linear regression indicates by how much the subject's estimate changes on average when the normative estimate changes by one unit. Many studies on human estimates, especially for probability judgments, have observed that the slope of the regression was less than 1 (Costello & Watts, 2014; Erev et al., 1994; Hilbert, 2012; Phillips & Edwards, 1966; Zhu et al., 2020). Consistent with these studies, we also observed in our study that the slope was less than 1, in both tasks (see Table S2 for descriptive and inferential statistics). In the literature, this phenomenon has been referred to as "conservatism bias" (Costello & Watts, 2014; Erev et al., 1994; Hilbert, 2012; Phillips & Edwards, 1966; Zhu et al., 2020), because a regression with a slope less than 1 predicts that, for a given level of normative estimate, the subject's estimate will be on average less close to the extremes (0 or 1, hence the 'conservatism' label), i.e. closer to 0.5, than the normative estimate. Here, we do not attach any particular mechanistic interpretation to the slope and treat it as a descriptive measure. For possible explanations of this phenomenon, see (Costello & Watts, 2014; Erev et al., 1994; Hilbert, 2012; Zhu et al., 2020).

Table S1. Pearson correlation coefficient between normative estimates and subjects' estimates.

Task	Mean	S.e.m.	Standard deviation	T-test against 0	
				t statistic	p value
Magnitude learning	0.96	0.01	0.09	106.20	2E-100
Probability learning	0.80	0.01	0.14	55.79	2E-74

Table S2. Slope of the linear regression between normative estimates and subjects' estimates.

Task	Mean	S.e.m.	Standard deviation	T-test against 0		T-test against 1	
				t statistic	p value	t statistic	p value
Magnitude learning	0.95	0.01	0.10	88.63	4E-93	4.79	6E-06
Probability learning	0.88	0.02	0.23	37.03	3E-58	4.96	3E-06

Decomposition of the mean squared error.

As presented in the main text, we performed a decomposition of the mean squared error between the subjects' estimates and the normative estimates to quantify the proportion of the error that was attributable to systematic biases in their estimates rather than to their variance (see Results).

We also conducted an additional analysis to investigate the bias: Since we observed a regression slope less than 1 consistent with a "conservatism bias" (Table S2), we investigated the extent to which such a conservatism bias could explain the subjects' bias. Specifically, we quantified the amount of bias explained by a linear regression model fitted to the subjects, which applies a linear transformation to the normative estimates, and models a conservatism bias when its slope is less than 1. We performed a linear regression between the normative estimates and the subjects' estimates averaged across the group, took the predictions of this regression as a model of the biased estimates, and then calculated the mean squared error obtained by replacing the normative estimates with the biased estimates. The proportion of the mean squared error that was reduced by using the biased estimates (i.e. the obtained reduction of the error in proportion to the original error) measures the amount of bias explained by the conservatism bias in subjects.

The full results of the decomposition (proportion of bias, variance, and of conservatism bias) are reported in Table S3 below.

Table S3. Decomposition of mean squared error between subjects' estimates and the normative estimates. MSE: mean squared error.

Task	Proportion of MSE due to bias	Proportion of MSE due to variance	Standard error of the proportion of bias/variance	Proportion of MSE explained by the conservatism bias	Standard error of the proportion of conservatism bias
Magnitude learning	23.48%	76.52%	1.12%	3.14%	0.63%
Probability learning	17.27%	82.73%	0.72%	3.23%	0.53%

Supplementary Text 2: Regression on subject's learning rate in the magnitude learning task performed as in a previous study.

For comparison with previous studies on magnitude learning, we additionally performed a regression analysis on the subject's learning rate as it was done in (McGuire et al., 2014), that is, without z-scoring regressors as we did in the regression reported in the main text, and replacing our prior uncertainty regressor by the RU*(1-CPP) regressor. We obtained regression weights similar to but slightly higher than those reported in (McGuire et al., 2014): The median and interquartile range of the regression weights in this analysis are 0.83 [0.56–0.92] and 0.51 [0.24–0.73] in our data for change-point probability and RU*(1-CPP) respectively (all two-tailed signed-rank $p<0.001$), vs. 0.53 [0.40–0.76] and 0.32 [0.11–0.44] in (McGuire et al., 2014).

Supplementary Text 3: Noisy delta-rule simulations

To examine the possibility that the learning rate adjustments observed in subjects could emerge from learning noise, we conducted simulations of a noisy delta rule model, with noise in the update similar to (Drugowitsch et al., 2016; Findling et al., 2019). The model is described by the following update equation:

$$v_t = v_{t-1} + \eta\,(x_t - v_{t-1}) + \varepsilon_t$$

where v_t is the model's estimate following observation x_t, η is the delta-rule parameter, and ε_t is the noise in the update, which is sampled from a zero-mean Gaussian distribution whose standard deviation corresponds to the noise level

We tested two variants for the noise level in the model: one (version a) where, as in (Drugowitsch et al., 2016), it is a constant parameter of the model, σ_ε, and another (version b) where, as in (Findling et al., 2019), it is scaled to the prediction error, with a scaling factor parameter ζ. Thus, the noise sampling is $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon)$ in version (a) and $\varepsilon_t \sim \mathcal{N}(0, \zeta\,|x_t - v_{t-1}|)$ in version (b).

To compute the parameter values, we leveraged the following properties of the model (E denotes the expectation, SD the standard deviation):

$\eta = E[(v_t - v_{t-1}) / (x_t - v_{t-1})]$, since $E[\varepsilon_t / (x_t - v_{t-1})] = 0$.

$\sigma_\varepsilon = SD[(v_t - v_{t-1}) - \eta\,(x_t - v_{t-1})]$, for version (a).

$\zeta = SD[((v_t - v_{t-1}) - \eta\,(x_t - v_{t-1})) / |x_t - v_{t-1}|]$, for version (b).

By computing the means and standard deviations described above across time and sequences, we obtained, for each subject and task, the parameter values that best match the subject's estimates (Fig. S5B).

For the results obtained with this model regarding learning rate adjustments, see Fig. S5C.

Supplementary Text 4: Correlations across subjects between the magnitude learning task and the probability learning task.

We correlated the regression weights obtained in the magnitude learning task with those obtained in the probability learning task across subjects (the weights were obtained using the same regression as in Fig. 5). The weight of change-point probability was not significantly correlated between the two tasks (r=-0.01, p=0.92), and that of prior uncertainty was weakly though significantly correlated (r=0.26, p=0.012) (partial Pearson correlation controlling for the individual's average update frequency, two-tailed p values). This is indeed due to differences between the two tasks: within each task, when performing the same correlation analysis on two halves of the data (even and odd sessions), we obtained strong correlations (these were, in the magnitude and probability task respectively, r=0.64 and 0.95 for change-point-probability, r=0.38 and 0.74 for prior uncertainty, all p<0.001, two-tailed). For comparison, another behavioral measure, the average update frequency of the subject, was more strongly correlated across subjects between the two tasks: r=0.63, p<0.001 (Pearson correlation, two-tailed).

Supplementary Figures

Continued on next page

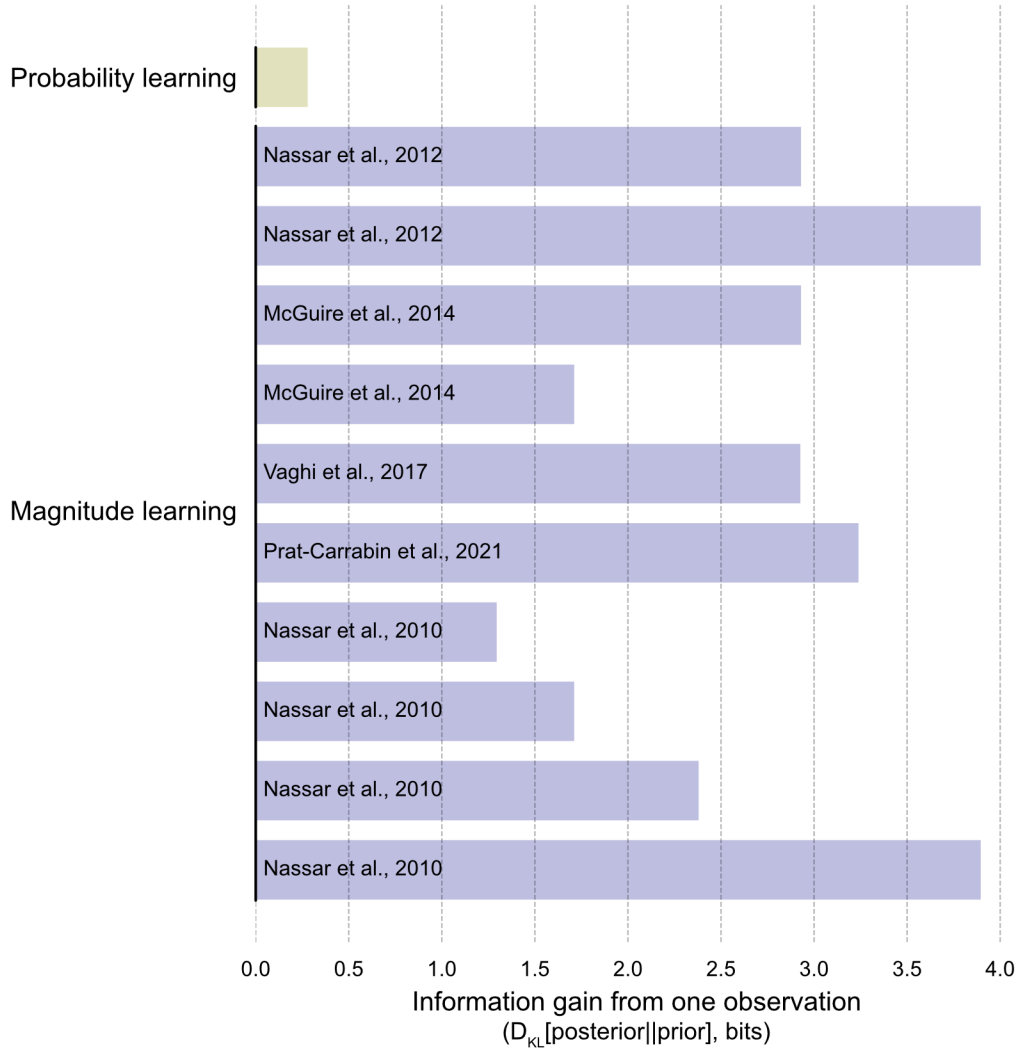


Fig. S1. Information gain provided by a single observation about the quantity to be learned in magnitude learning and probability learning. We computed the information gain as the KL divergence between the prior (uniform) distribution before having received any observation, and the posterior distribution about the underlying quantity after having received the observation, using the prior as reference distribution (i.e. D_{KL}[posterior||prior]), on average over the possible observations. The posterior is obtained from the prior and the likelihood function relating the observation to the underlying quantity using Bayes rule. In probability learning, the information gain is minimal. This is due to the binary nature of the observation. In magnitude learning, the information gain is larger because the observation is quantitative and typically fairly representative of the underlying magnitude. Although the latter depends on the experimenter's choice of standard deviation with which observations are generated, we computed the information gain for numerous experiments previously conducted and each condition of these experiments, and as shown above, in all cases it was substantially higher than that obtained in probability learning (McGuire et al., 2014; Nassar et al., 2010, 2012; Prat-Carrabin et al., 2021; Vaghi et al., 2017).

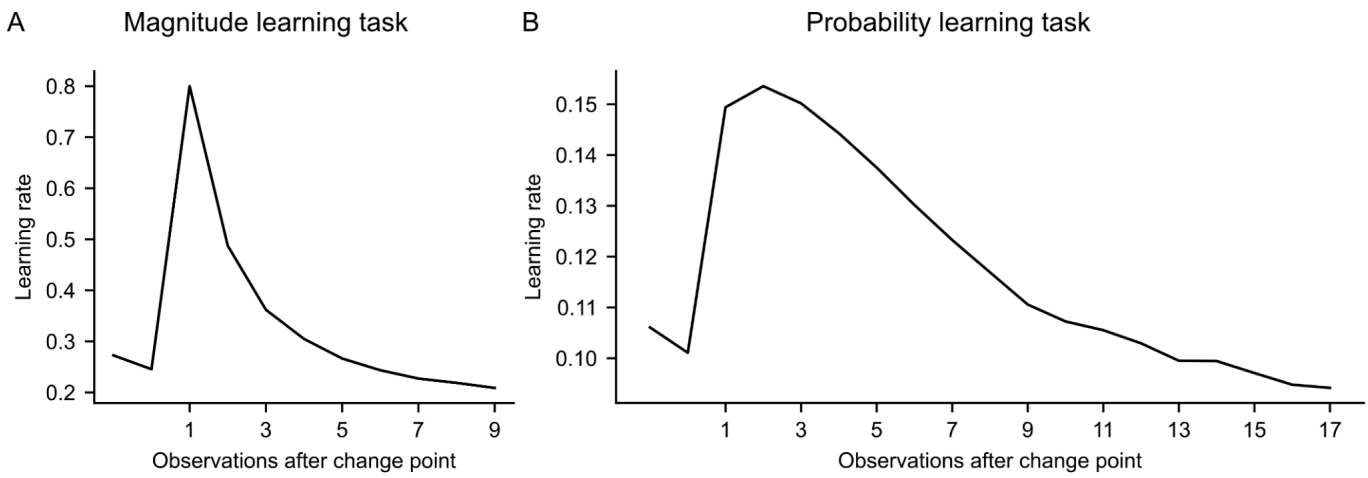


Fig. S2. Dynamics of the normative model's learning rate after a change point, in the magnitude (A) and probability (B) learning tasks. The plots were obtained as in Fig. 2, but rather than using the subjects' learning rate, we used the normative model's learning rate instead, which we obtained by running the normative model on the same sequences as the subject.

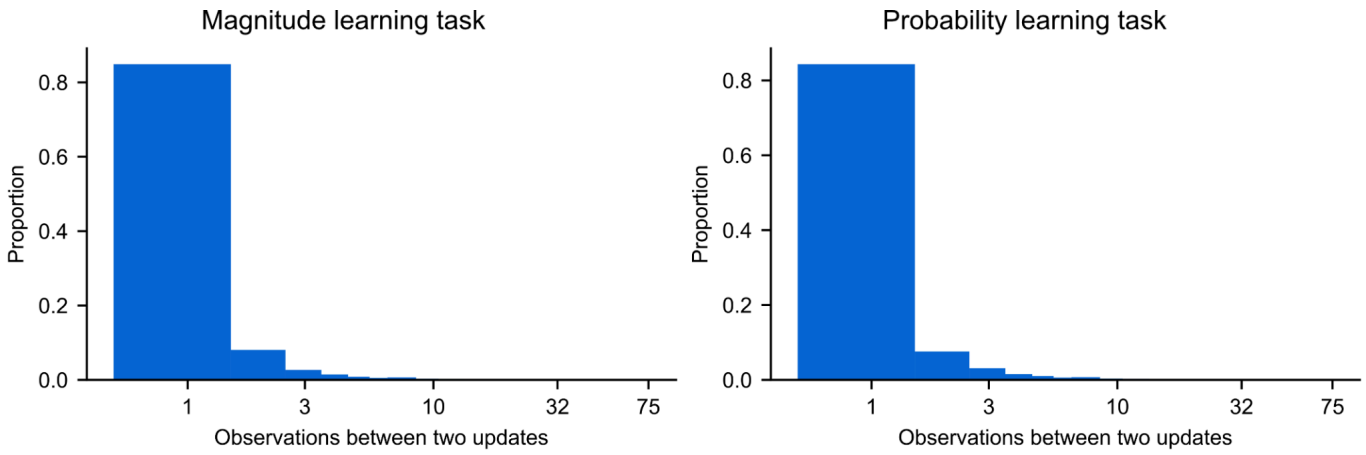


Fig. S3. Distribution of the number of observations elapsed between two report updates made by subjects. A log scale was used for the number of observations as in (Gallistel et al., 2014) for comparison (the equivalent distribution in Gallistel et al. is shown in their Fig. 11). In contrast to (Gallistel et al., 2014), in our study, updates were made on each observation most of the time (84% in the above distribution; mean of the distribution: 1.4 observation).

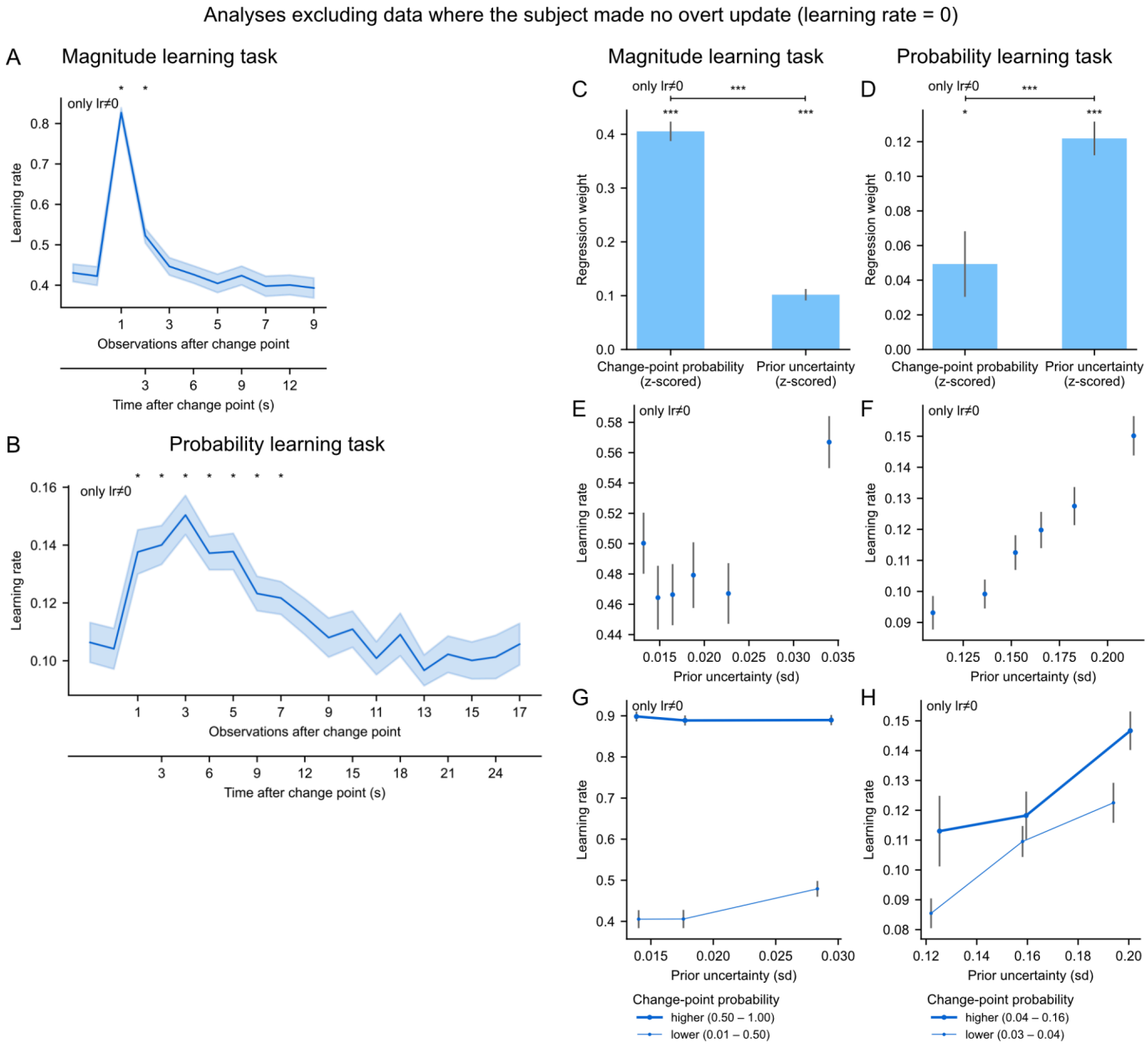


Fig. S4. The main results are similar and remain significant when excluding all data where the subject did not make an overt update. After excluding all data points where this was the case (i.e. learning rate = 0), we performed the same analyses as in previous figures and obtained the above plots: (A and B) Equivalent to Fig. 2 A and B; (C–H) Equivalent to Fig. 5 (A–F). Stars denote statistical significance as in the main figures (see legends of those figures for further details).

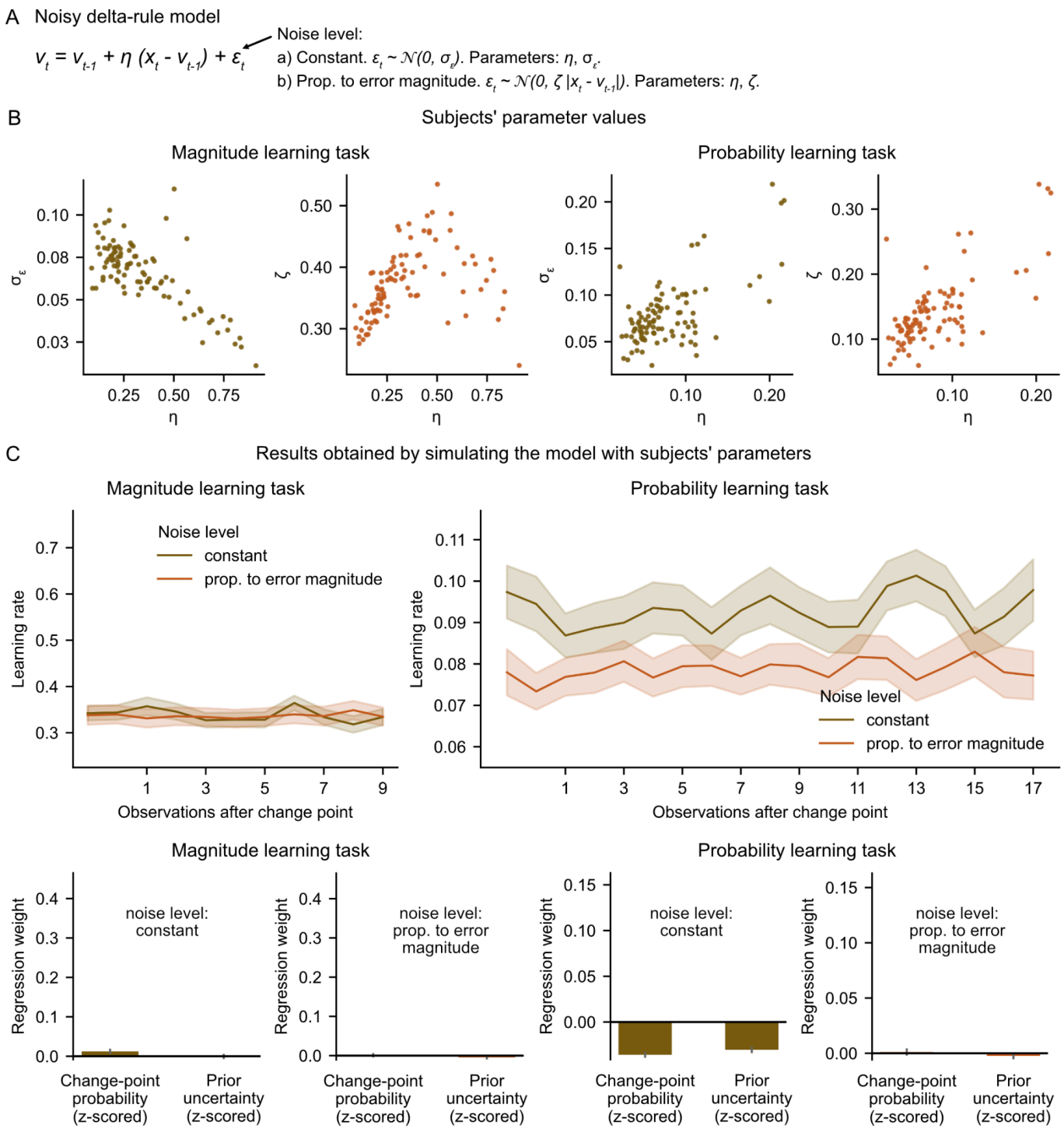


Fig. S5. The subjects' dynamic adjustments of the learning rate are not explained by learning noise. (A) Model of a delta-rule with learning noise. A noise sample is injected at each update of the model, otherwise governed by a delta-rule with parameter η. Two versions were tested for the noise level: (a) constant (parameter σ_ε), (b) scaled to the magnitude of the prediction error (scaling factor parameter ζ). (B) Values of the model parameters for each subject, for each version of the model and each task. Each dot represents one subject. (C) Results obtained by simulating the model with the subject's parameters on the subject's sequences and performing the same learning rate analyses as those reported in the main results for subjects. Top plots are the results for the analysis corresponding to Fig. 2, bottom plots to Fig. 5.

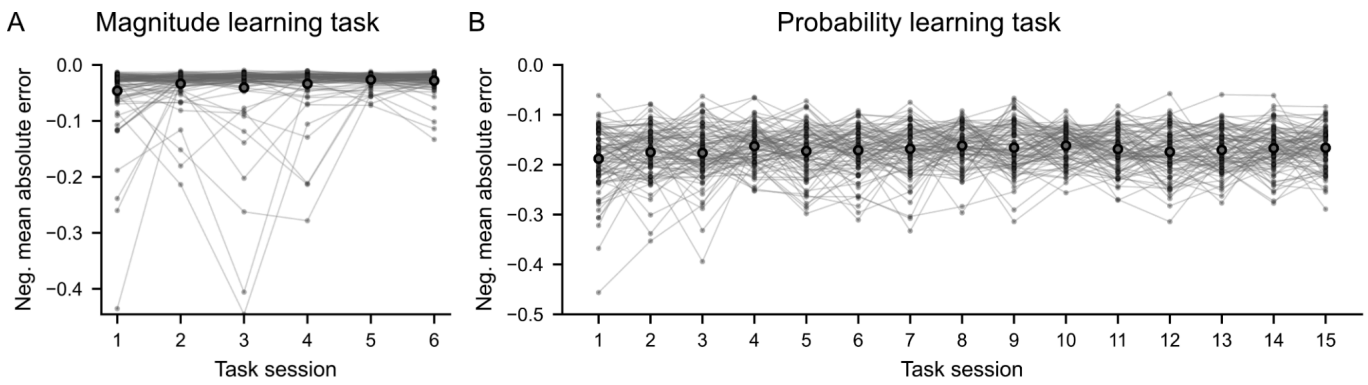


Fig. S6. Subjects' performance was stable over the course of the task. Performance is measured by the accuracy of the estimates, quantified by the mean absolute error between the subject's estimate and the true value of the hidden quantity (the negative of the error was used so that higher values correspond to higher performance). Thin dots and lines connecting them each denote one subject; large circles denote the mean across subjects.

Chapter III: Article 2, Theoretical study (Foucault & Meyniel, 2021, published in eLife)

eLife online version: https://doi.org/10.7554/eLife.71801

RESEARCH ARTICLE

Gated recurrence enables simple and accurate sequence prediction in stochastic, changing, and structured environments

Cédric Foucault[1,2], Florent Meyniel[1]*

[1]Cognitive Neuroimaging Unit, INSERM, CEA, Université Paris-Saclay, NeuroSpin center, Gif sur Yvette, France; [2]Sorbonne Université, Collège Doctoral, Paris, France

Abstract From decision making to perception to language, predicting what is coming next is crucial. It is also challenging in stochastic, changing, and structured environments; yet the brain makes accurate predictions in many situations. What computational architecture could enable this feat? Bayesian inference makes optimal predictions but is prohibitively difficult to compute. Here, we show that a specific recurrent neural network architecture enables simple and accurate solutions in several environments. This architecture relies on three mechanisms: gating, lateral connections, and recurrent weight training. Like the optimal solution and the human brain, such networks develop internal representations of their changing environment (including estimates of the environment's latent variables and the precision of these estimates), leverage multiple levels of latent structure, and adapt their effective learning rate to changes without changing their connection weights. Being ubiquitous in the brain, gated recurrence could therefore serve as a generic building block to predict in real-life environments.

***For correspondence:** florent.meyniel@cea.fr

Competing interest: The authors declare that no competing interests exist.

Funding: See page 27

Preprinted: 03 May 2021
Received: 30 June 2021
Accepted: 01 December 2021
Published: 02 December 2021

Reviewing Editor: Srdjan Ostojic, Ecole Normale Superieure Paris, France

Editor's evaluation

There has been a longstanding interest in developing normative models of how humans handle latent information in stochastic and volatile environments. This study examines recurrent neural network models trained on sequence-prediction tasks analogous to those used in human cognitive studies. The results demonstrate that such models lead to highly accurate predictions for challenging sequences in which the statistics are non-stationary and change at random times. These novel and remarkable results open up new avenues for cognitive modelling.

Introduction

Being able to correctly predict what is coming next is advantageous: it enables better decisions (***Dolan and Dayan, 2013***; ***Sutton and Barto, 1998***), a more accurate perception of our world, and faster reactions (***de Lange et al., 2018***; ***Dehaene et al., 2015***; ***Saffran et al., 1996***; ***Sherman et al., 2020***; ***Summerfield and de Lange, 2014***). In many situations, predictions are informed by a sequence of past observations. In that case, the prediction process formally corresponds to a statistical inference that uses past observations to estimate latent variables of the environment (e.g. the probability of a stimulus) that then serve to predict what is likely to be observed next. Specific features of real-life environments make this inference a challenge: they are often partly random, changing, and structured in different ways. Yet, in many situations, the brain is able to overcome these challenges and shows several aspects of the optimal solution (***Dehaene et al., 2015***; ***Dolan and Dayan, 2013***; ***Gallistel***

et al., 2014; *Summerfield and de Lange, 2014*). Here, we aim to identify the computational mechanisms that could enable the brain to exhibit these aspects of optimality in these environments.

We start by unpacking two specific challenges which arise in real-life environments. First, the joint presence of randomness and changes (i.e. the non-stationarity of the stochastic process generating the observations) poses a well-known tension between stability and flexibility (*Behrens et al., 2007*; *Soltani and Izquierdo, 2019*; *Sutton, 1992*). Randomness in observations requires integrating information over time to derive a stable estimate. However, when a change in the estimated variable is suspected, it is better to limit the integration of past observations to update the estimate more quickly. The prediction should thus be adaptive, that is, dynamically adjusted to promote flexibility in the face of changes and stability otherwise. Past studies have shown that the brain does so in many contexts: perception (*Fairhall et al., 2001*; *Wark et al., 2009*), homeostatic regulation (*Pezzulo et al., 2015*; *Sterling, 2004*), sensorimotor control (*Berniker and Kording, 2008*; *Wolpert et al., 1995*), and reinforcement learning (*Behrens et al., 2007*; *Iglesias et al., 2013*; *Soltani and Izquierdo, 2019*; *Sutton and Barto, 1998*).

Second, the structure of our environment can involve complex relationships. For instance, the sentence beginnings "what science can do for you is..." and "what you can do for science is..." call for different endings even though they contain the same words, illustrating that prediction takes into account the ordering of observations. Such structures appear not only in human language but also in animal communication (*Dehaene et al., 2015*; *Hauser et al., 2001*; *Robinson, 1979*; *Rose et al., 2004*), and all kinds of stimulus-stimulus and stimulus-action associations in the world (*Saffran et al., 1996*; *Schapiro et al., 2013*; *Soltani and Izquierdo, 2019*; *Sutton and Barto, 1998*). Such a structure is often latent (i.e. not directly observable) and it governs the relationship between observations (e.g. words forming a sentence, stimulus-action associations). These relationships must be leveraged by the prediction, making it more difficult to compute.

In sum, the randomness, changes, and latent structure of real-life environments pose two major challenges: that of adapting to changes and that of leveraging the latent structure. Two commonly used approaches offer different solutions to these challenges. The Bayesian approach allows to derive statistically optimal predictions for a given environment knowing its underlying generative model. This optimal solution is a useful benchmark and has some descriptive validity since, in some contexts, organisms behave close to optimally (*Ma and Jazayeri, 2014*; *Tauber et al., 2017*) or exhibit several qualitative aspects of the optimal solution (*Behrens et al., 2007*; *Heilbron and Meyniel, 2019*; *Meyniel et al., 2015*). However, a specific Bayes-optimal solution only applies to a specific generative model (or class of models [*Tenenbaum et al., 2011*]). This mathematical solution also does not in general lead to an algorithm of reasonable complexity (*Cooper, 1990*; *Dagum and Luby, 1993*). Bayesian inference therefore says little about the algorithms that the brain could use, and the biological basis of those computations remains mostly unknown with only a few proposals highly debated (*Fiser et al., 2010*; *Ma et al., 2006*; *Sahani and Dayan, 2003*).

Opposite to the Bayes-optimal approach is the heuristics approach: solutions that are easy to compute and accurate in specific environments (*Todd and Gigerenzer, 2000*). However, heuristics lack generality: their performance can be quite poor outside the environment that suits them. In addition, although simple, their biological implementation often remains unknown (besides the delta-rule [*Eshel et al., 2013*; *Rescorla and Wagner, 1972*; *Schultz et al., 1997*]).

Those two approaches leave open the following questions: Is there a general, biologically feasible architecture that enables, in different environments, solutions that are simple, effective, and that reproduce the qualitative aspects of optimal prediction observed in organisms? If so, what are its essential mechanistic elements?

Our approach stands in contrast with the elegant closed-form but intractable mathematical solutions offered by Bayesian inference, and the simple but specialized algorithms offered by heuristics. Instead, we look for general mechanisms under the constraints of feasibility and simplicity. We used recurrent neural networks because they can offer a generic, biologically feasible architecture able to realize different prediction algorithms (see *LeCun et al., 2015*; *Saxe et al., 2021* and Discussion). We used small network sizes in order to produce simple (i.e. low-complexity, memory-bounded) solutions. We tested their generality using different environments. To determine the simplest architecture sufficient for effective solutions and derive mechanistic insights, we considered different architectures that varied in size and mechanisms. For each one, we instantiated several networks and

trained them to approach their best possible prediction algorithm in a given environment. We treated the training procedure as a methodological step without claiming it to be biologically plausible. To provide interpretability, we inspected the networks' internal model and representations, and tested specific optimal aspects of their behavior—previously reported in humans (***Heilbron and Meyniel, 2019***; ***Meyniel et al., 2015***; ***Nassar et al., 2010***; ***Nassar et al., 2012***)—which demonstrate the ability to adapt to changes and leverage the latent structure of the environment.

Results

The framework: sequence prediction and network architectures

All our analyses confront simulated agents with the same general problem: sequence prediction. It consists in predicting, at each time step in a sequence where one time step represents one observation, the probability distribution over the value of the next observation given the previous observations (here we used binary observations coded as '0' and '1') (***Figure 1a***). The environment generates the sequence, and the agent's goal is to make the most accurate predictions possible in this environment. Below, we introduce three environments. All of them are stochastic (observations are governed by latent probabilities) and changing (these latent probabilities change across time), and thus require dynamically adapting the stability-flexibility tradeoff. They also feature increasing levels of latent structure that must be leveraged, making the computation of predictions more complex.

How do agents learn to make predictions that fit a particular environment? In real life, agents often do not benefit from any external supervision and must rely only on the observations. To do so, they can take advantage of an intrinsic error signal that measures the discrepancy between their prediction and the actual value observed at the next time step. We adopted this learning paradigm (often called unsupervised, self-supervised, or predictive learning in machine learning [***Elman, 1991***; ***LeCun, 2016***]) to train our agents in silico. We trained the agents by exposing them to sequences generated by a given environment and letting them adjust their parameters to improve their prediction (see Materials and methods).

During testing, we kept the parameters of the trained agents frozen, exposed them to new sequences, and performed targeted analyses to probe whether they exhibit specific capabilities and better understand how they solve the problem.

Our investigation focuses on a particular class of agent architectures known as recurrent neural networks. These are well suited for sequence prediction because recurrence allows to process inputs sequentially while carrying information over time in recurrent activity. The network architectures we used all followed the same three-layer template, consisting of one input unit whose activity codes for the current observation, one output unit whose activity codes for the prediction about the next observation, and a number of recurrent units that are fed by the input unit and project to the output unit (***Figure 1b***). All architectures had self-recurrent connections.

We identified three mechanisms of recurrent neural network architectures that endow a network with specific computational properties which have proven advantageous in our environments (***Figure 1c***). One mechanism is gating, which allows for multiplicative interactions between the activities of units. A second mechanism is lateral connectivity, which allows the activities of different recurrent units to interact with each other. A third mechanism is the training of recurrent connection weights, which allows the dynamics of recurrent activities to be adjusted to the training environment.

To get mechanistic insight, we compared an architecture that included all three mechanisms, to alternative architectures that were deprived of one of the three mechanisms but maintained the other two (***Figure 1d***; see Materials and methods for equations). Here, we call an architecture with all three mechanisms 'gated recurrent', and the particular architecture we used is known as GRU (***Cho et al., 2014***; ***Chung et al., 2014***). When deprived of gating, multiplicative interactions between activities are removed, and the architecture reduces to that of a vanilla recurrent neural network also known as the Elman network (***Elman, 1990***). When deprived of lateral connections, the recurrent units become independent of each other, thus each recurrent unit acts as a temporal filter on the input observations (with possibly time-varying filter weights thanks to gating). When deprived of recurrent weight training, the recurrent activity dynamics become independent of the environment and the only parameters that can be trained are those of the output unit; this architecture is thus one form of reservoir computing (***Tanaka et al., 2019***). In the results below, unless otherwise stated, the networks all had

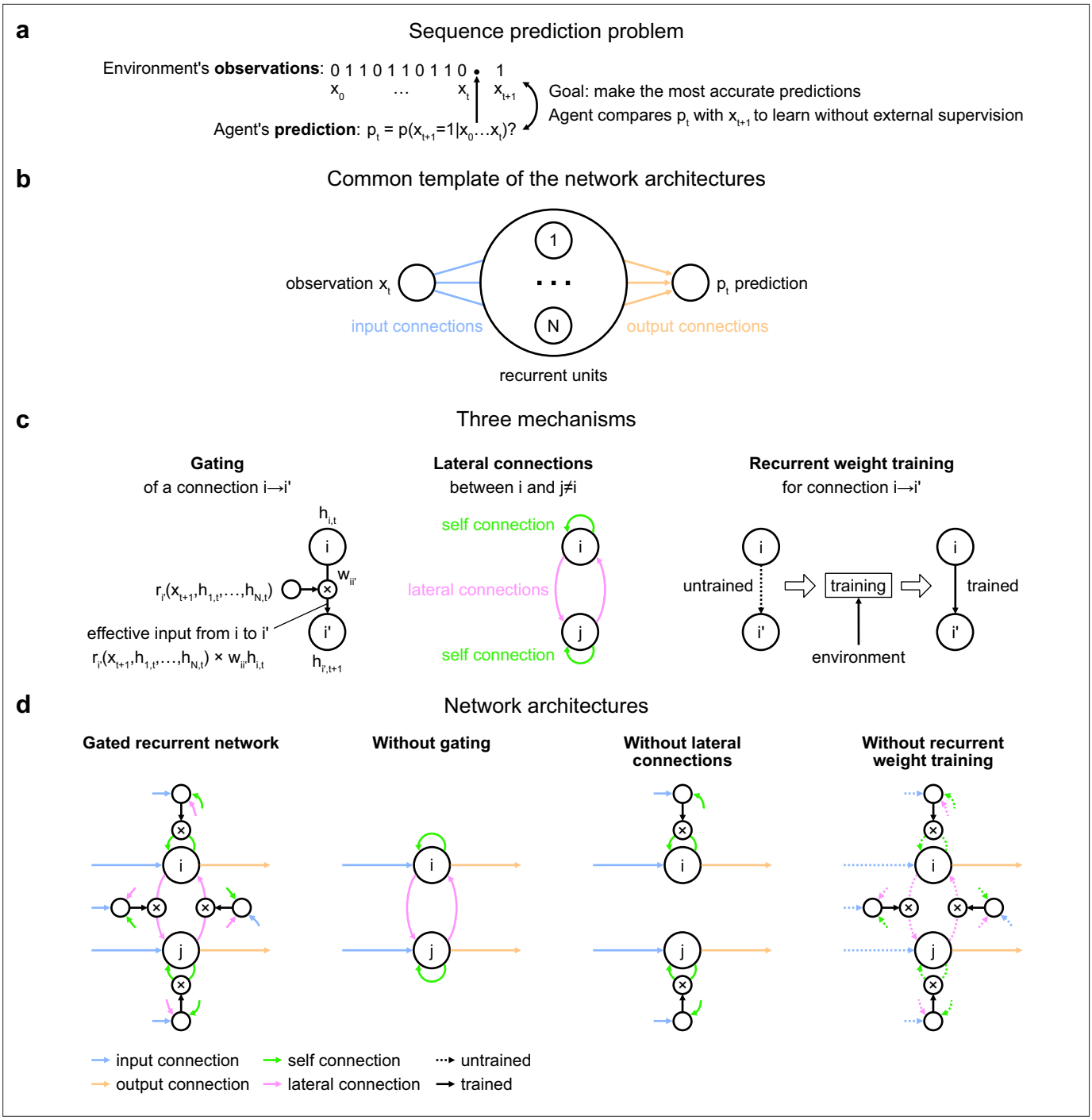


Figure 1. Problem to solve and network architectures. (**a**) Sequence prediction problem. At each time step t, the environment generates one binary observation x_t. The agent receives it and returns a prediction p_t: its estimate of the probability that the next observation will be one given the observations collected so far. The agent's goal is to make the most accurate predictions possible. The agent can measure its accuracy by comparing its prediction p_t with the actual value observed at the next time step x_{t+1}, allowing it to learn from the observations without any external supervision. (**b**) Common three-layer template of the recurrent neural network architectures. Input connections transmit the observation to the recurrent units and output connections allow the prediction to be read from the recurrent units. (**c**) Three key mechanisms of recurrent neural network architectures. Gating allows for multiplicative interaction between activities. Lateral connections allow the activities of different recurrent units i and j to interact. Recurrent weight training allows the connection weights of recurrent units to be adjusted to the training environment. i' may be equal to i. (**d**) The gated recurrent architecture includes all three mechanisms: gating, lateral connections, and recurrent weight training. Each alternative architecture includes all but one of the three mechanisms.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure 1 continued on next page

Figure 1 continued

Figure supplement 1. Graphical model of the generative process of each environment.

11 recurrent units (the smallest network size beyond which the gated recurrent network showed no substantial increase in performance in any of the environments), but the results across architectures are robust to this choice of network size (see the last section of the Results).

Performance in the face of changes in latent probabilities

We designed a first environment to investigate the ability to handle changes in a latent probability (***Figure 2a***; see ***Figure 1—figure supplement 1*** for a graphical model). In this environment we used the simplest kind of latent probability: p(1), the probability of occurrence (or base rate) of the observation being 1 (note that p(0) = 1−p(1)), here called 'unigram probability'. The unigram probability suddenly changed from one value to another at so-called 'change points', which could occur at any time, randomly with a given fixed probability.

This environment, here called 'changing unigram environment', corresponds for instance to a simple oddball task (***Aston-Jones et al., 1997***; ***Kaliukhovich and Vogels, 2014***; ***Ulanovsky et al., 2004***), or the probabilistic delivery of a reward with abrupt changes in reward probabilities (***Behrens et al., 2007***; ***Vinckier et al., 2016***). In such an environment, predicting accurately is difficult due to the stability-flexibility tradeoff induced by the stochastic nature of the observations (governed by the unigram probability) and the possibility of a change point at any moment.

To assess the networks' prediction accuracy, we compared the networks with the optimal agent for this specific environment, that is, the optimal solution to the prediction problem determined using Bayesian inference. This optimal solution knows the environment's underlying generative process and uses it to compute, via Bayes' rule, the probability distribution over the possible values of the latent probability given the past observation sequence, $p(p^{env}_{t+1}|x_0, ..., x_t)$ known as the posterior distribution. It then outputs as prediction the mean of this distribution. (For details see Materials and methods and ***Heilbron and Meyniel, 2019***).

We also compared the networks to two types of heuristics which perform very well in this environment: the classic 'delta-rule' heuristic (***Rescorla and Wagner, 1972***; ***Sutton and Barto, 1998***) and the more accurate 'leaky' heuristic (***Gijsen et al., 2021***; ***Heilbron and Meyniel, 2019***; ***Meyniel et al.,***

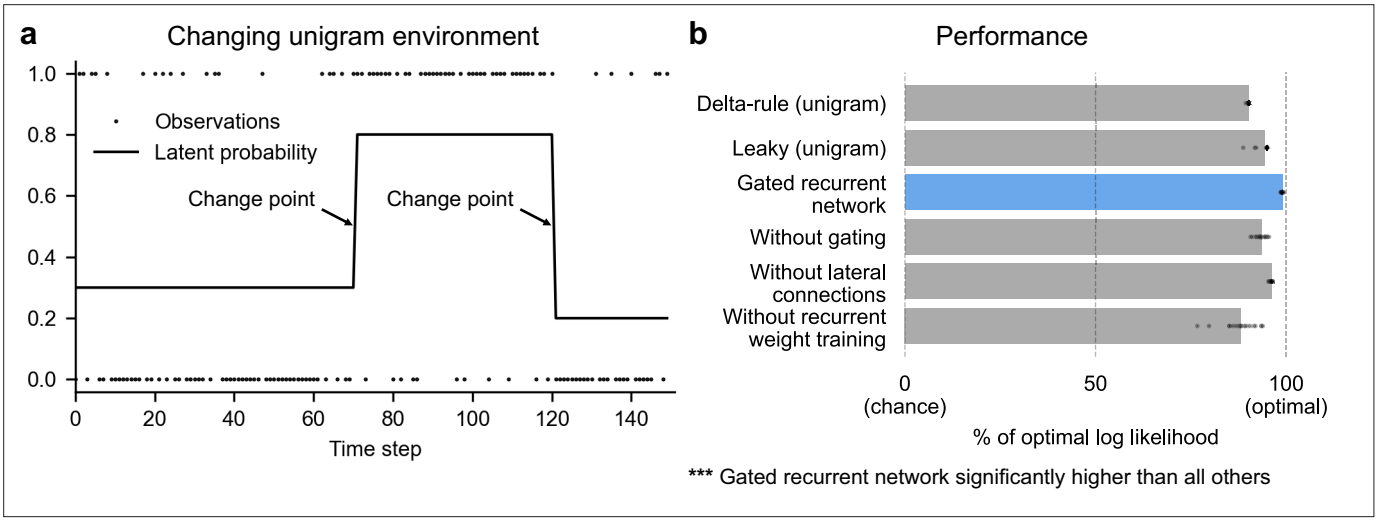


Figure 2. Gated recurrent networks perform quasi-optimally in the face of changes in latent probabilities. (**a**) Sample sequence of observations (dots) and latent unigram probability (line) generated in the changing unigram environment. At each time step, a binary observation is randomly generated based on the latent unigram probability, and a change point can occur with a fixed probability, suddenly changing the unigram probability to a new value uniformly drawn in [0,1]. (**b**) Prediction performance in the changing unigram environment. For each type of agent, 20 trained agents (trained with different random seeds) were tested (dots: agents; bars: average). Their prediction performance was measured as the % of optimal log likelihood (0% being chance performance and 100% optimal performance, see ***Equation 1*** for the log likelihood) and averaged over observations and sequences. The gated recurrent network significantly outperformed every other type of agent ($p < 0.001$, two-tailed two independent samples t-test with Welch's correction for unequal variances).

2016; *Yu and Cohen, 2008*) (see Materials and methods for details). To test the statistical reliability of our conclusions, we trained separately 20 agents of each type (each type of network and each type of heuristic).

We found that even with as few as 11 units, the gated recurrent networks performed quasi-optimally. Their prediction performance was 99% of optimal (CI±0.1%), 0% corresponding to chance level (*Figure 2b*). Being only 1% short of optimal, the gated recurrent networks outperformed the delta rule and leaky agents, which performed 10 times and 5 times further from optimal, respectively (*Figure 2b*).

For mechanistic insight, we tested the alternative architectures deprived of one mechanism. Without either gating, lateral connections, or recurrent weight training, the average performance was respectively 6 times, 4 times, and 12 times further from optimal (*Figure 2b*), that is, the level of a leaky agent or worse. The drops in performance remain similar when considering only the best network of each architecture instead of the average performance (*Figure 2b*, compare rightmost dots across rows).

These results show that small gated recurrent networks can achieve quasi-optimal predictions and that the removal of one of the mechanisms of the gated recurrent architecture results in a systematic drop in performance.

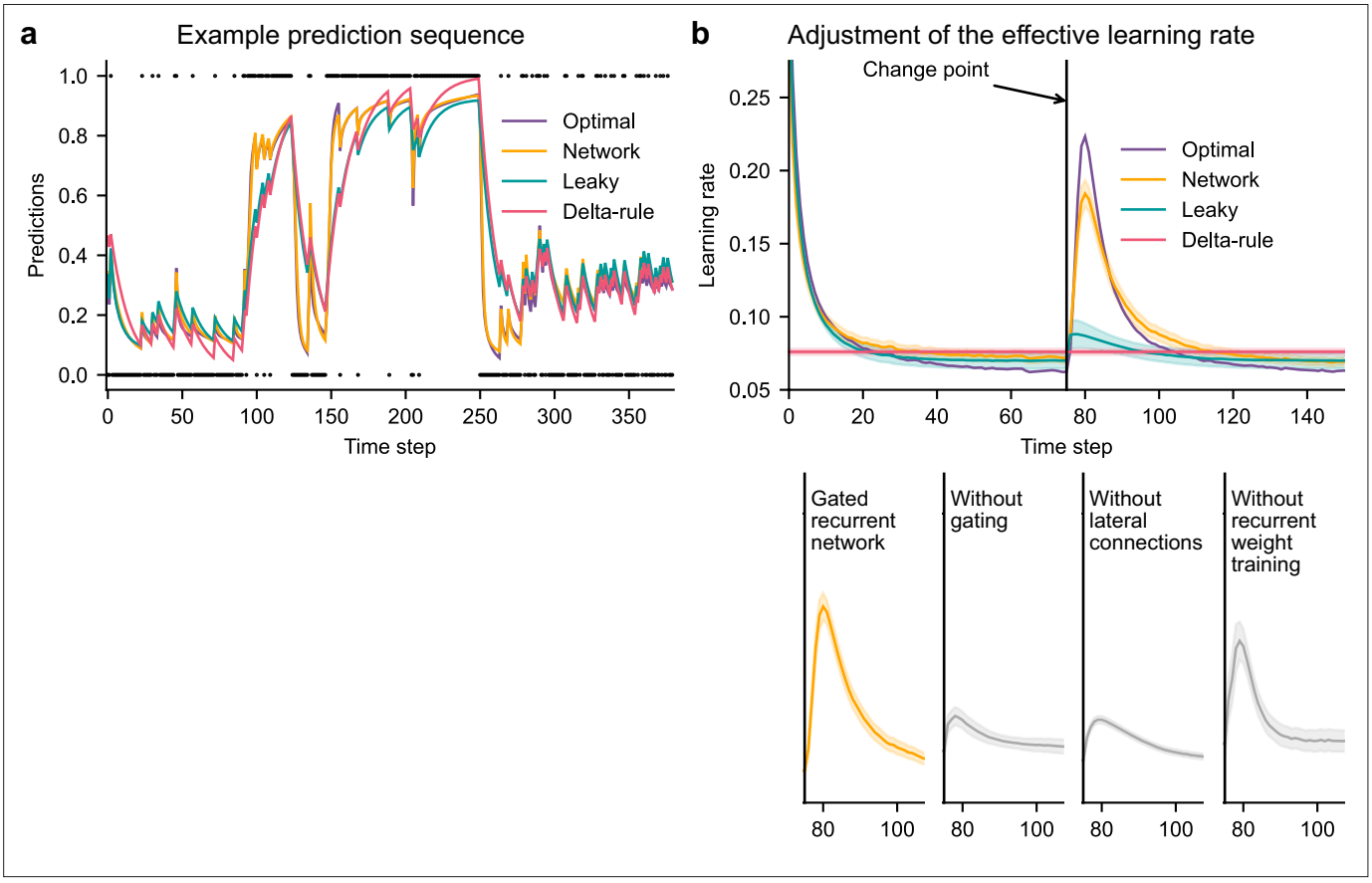


Figure 3. Gated recurrent but not alternative networks adjust their moment-by-moment effective learning rate around changes like the optimal agent. (**a**) Example prediction sequence illustrating the prediction updates of different types of agents. Within each type of agent, the agent (out of 20) yielding median performance in *Figure 2b* was selected for illustration purposes. Dots are observations, lines are predictions. (**b**) Moment-by-moment effective learning rate of each type of agent. 20 trained agents of each type were tested on 10,000 sequences whose change points were locked at the same time steps, for illustration purposes. The moment-by-moment effective learning rate was measured as the ratio of prediction update to prediction error (see Materials and methods, *Equation 2*), and averaged over sequences. Lines and bands show the mean and the 95% confidence interval of the mean.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Attunement of the effective learning rate to the change point probabilities.

Adaptation to changes through the adjustment of the effective learning rate

In a changing environment, the ability to adapt to changes is key. Networks exposed to more changing environments during training updated their predictions more overall during testing, similarly to the optimal agent (see ***Figure 3—figure supplement 1***) and, to some extent, humans (***Behrens et al., 2007***, Figure 2e; ***Findling et al., 2021***, Figure 4c). At a finer timescale, the moment-by-moment updating of the predictions also showed sensible dynamics around change points.

Figure 3a illustrates a key difference in behavior between, on the one hand, the optimal agent and the gated recurrent network, and on the other hand, the heuristic agents: the dynamics of their update differ. This difference is particularly noticeable when recent observations suggest that a change point has just occurred: the optimal agent quickly updates the prediction by giving more weight to the new observations; the gated recurrent network behaves the same but not the heuristic agents. We formally tested this dynamic updating around change points by measuring the moment-by-moment effective learning rate, which normalizes the amount of update in the prediction by the prediction error (i.e. the difference between the previous prediction and the actual observation; see Materials and methods, ***Equation 2***).

Gated recurrent networks turned out to adjust their moment-by-moment effective learning rate as the optimal agent did, showing the same characteristic peaks, at the same time and with almost the same amplitude (***Figure 3b***, top plot). By contrast, the effective learning rate of the delta-rule agents was (by construction) constant, and that of the leaky agents changed only marginally.

When one of the mechanisms of the gated recurrence was taken out, the networks' ability to adjust their effective learning rate was greatly degraded (but not entirely removed) (***Figure 3b***, bottom plots). Without gating, without lateral connections, or without recurrent weight training, the amplitude was lower (showing both a lower peak value and a higher baseline value), and the peak occurred earlier.

This shows that gated recurrent networks can reproduce a key aspect of optimal behavior: the ability to adapt the update of their prediction to change points, which is lacking in heuristic agents and alternative networks.

Internal representation of precision and dynamic interaction with the prediction

Beyond behavior, we sought to determine whether a network's ability to adapt to changes relied on idiosyncratic computations or followed the more general principle of precision-weighting derived from probability theory. According to this principle, the precision of the current prediction (calculated in the optimal agent as the negative logarithm of the standard deviation of the posterior distribution over the latent probability, see ***Equation 3*** in Materials and methods) should influence the weight of the current prediction relative to the next observation in the updating process: for a given prediction error, the lower the precision, the higher the subsequent effective learning rate. This precision-weighting principle results in an automatic adjustment of the effective learning rate in response to a change, because the precision of the prediction decreases when a change is suspected.

In line with this principle, human participants can estimate not only the prediction but also its precision as estimated by the optimal agent (***Boldt et al., 2019***, Figure 2; ***Meyniel et al., 2015***, Figure 4B), and this precision indeed relates to the participants' effective learning rate (***McGuire et al., 2014***, Figure 2C and S1A; ***Nassar et al., 2010***, Figure 4C and 3B; ***Nassar et al., 2012***, Figure 5 and 7c,).

We tested whether a network could represent this optimal precision too, by trying to linearly read it from the network's recurrent activity (***Figure 4a***). Note that the networks were trained only to maximize prediction accuracy (not to estimate precision). Yet, in gated recurrent networks, we found that the read precision on left-out data was highly accurate (***Figure 4a***, left plot: the median Pearson correlation with the optimal precision is 0.82), and correlated with their subsequent effective learning rate as in the optimal agent (***Figure 4a***, right plot: the median correlation for gated recurrent networks is –0.79; for comparison, it is –0.88 for the optimal agent).

To better understand how precision information is represented and how it interacts with the prediction dynamically in the network activity, we plotted the dynamics of the network activity in the subspace spanned by the prediction and precision vectors (***Figure 4b***). Such visualization captures both the temporal dynamics and the relationships between the variables represented in the network,

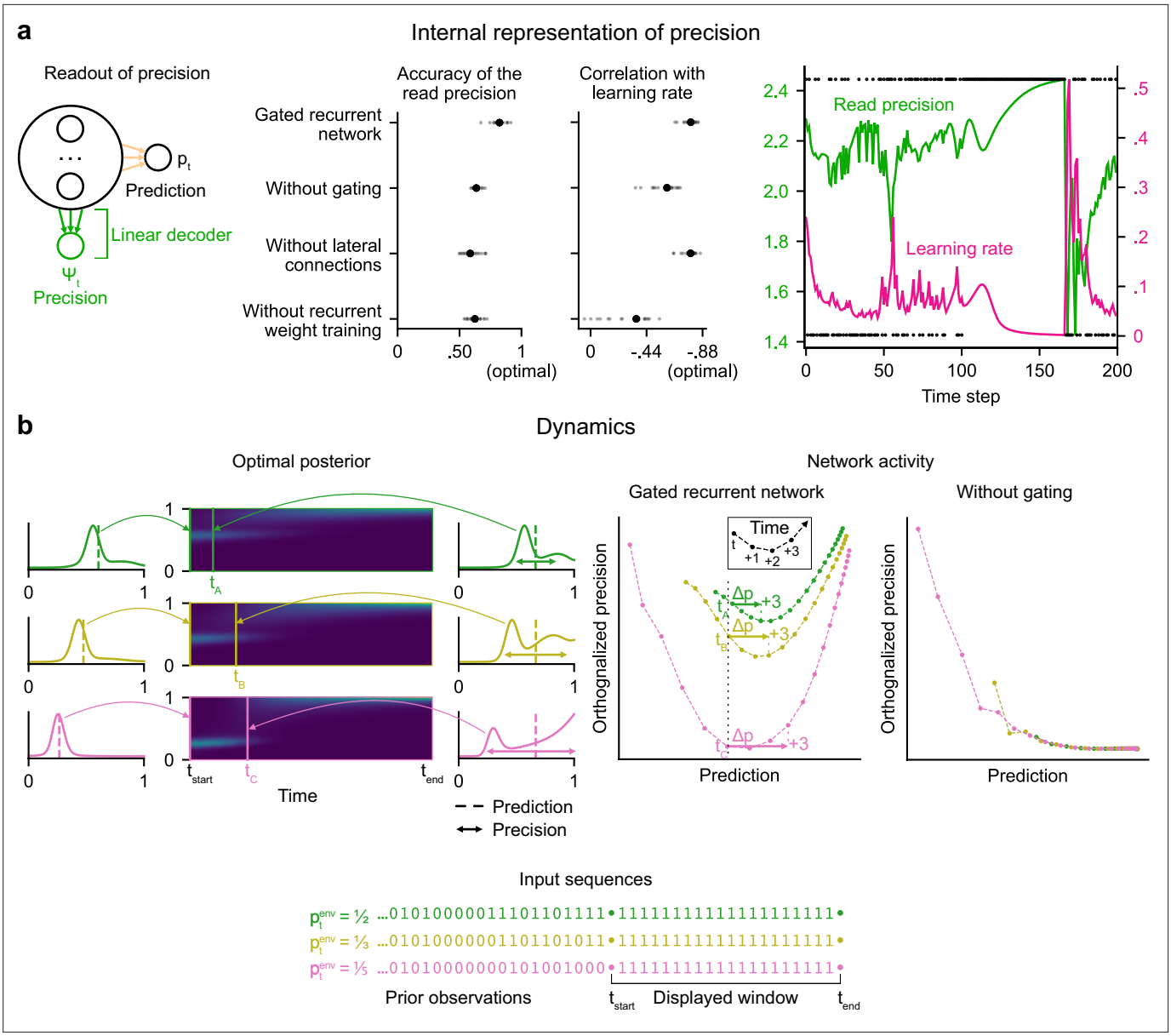


Figure 4. Gated recurrent networks have an internal representation of the precision of their estimate that dynamically interacts with the prediction following the precision-weighting principle. (**a**) Left to right: Schematic of the readout of precision from the recurrent activity of a network (obtained by fitting a multiple linear regression from the recurrent activity to the log precision of the optimal posterior distribution); Accuracy of the read precision (calculated as its Pearson correlation with the optimal precision); Pearson correlation between the read precision and the network's subsequent effective learning rate (the optimal value was calculated from the optimal agent's own precision and learning rate); Example sequence illustrating their anti-correlation in the gated recurrent network. In both dot plots, large and small dots show the median and individual values, respectively. (**b**) Dynamics of the optimal posterior (left) and the network activity (right) in three sequences (green, yellow, and pink). The displayed dynamics are responses to a streak of 1 s after different sequences of observations (with different generative probabilities as shown at the bottom). The optimal posterior distribution is plotted as a color map over time (dark blue and light green correspond to low and high probability densities, respectively) and as a line plot at two times: on the left, the time t_{start} just before the streak of 1s, and on the right, a time $t_A/t_B/t_C$ when the prediction (i.e. mean) is approximately equal in all three cases; note that the precision differs. The network activity was projected onto the two-dimensional subspace spanned by the prediction and precision vectors (for the visualization, the precision axis was orthogonalized with respect to the prediction axis). In the gated recurrent network, the arrow Δp shows the update to the prediction performed in the next three time steps starting at the time $t_A/t_B/t_C$ defined from the optimal posterior. Like the optimal posterior and unlike the network without gating, the gated recurrent network represents different levels of precision at an equal prediction, and the lower the precision, the higher the subsequent update to the prediction—a principle called precision-weighting. In all example plots (**a–b**), the displayed network is the one of the 20 that yielded the median read precision accuracy.

and has helped understand network computations in other works (***Mante et al., 2013***; ***Sohn et al., 2019***). Here, two observations can be made.

First, in the gated recurrent network (***Figure 4b***, second plot from the right), the trajectories are well separated along the precision axis (for the same prediction, the network can represent multiple precisions), meaning that the representation of precision is not reducible to the prediction. By contrast, in the network without gating (***Figure 4b***, rightmost plot), these trajectories highly overlap, which indicates that the representation of precision and prediction are mutually dependent. To measure this dependence, we computed the mutual information between the read precision and the prediction of the network, and it turned out to be very high in the network without gating (median MI = 5.2) compared to the gated recurrent network (median MI = 0.7) and the optimal agent (median MI = 0.6) (without lateral connections, median MI = 1.3; without recurrent weight training, median MI = 1.9), confirming that gating is important to separate the precision from the prediction.

Second, in the gated recurrent network, the precision interacts dynamically with the prediction in a manner consistent with the precision-weighting principle: for a given prediction, the lower the precision, the larger the subsequent updates to the prediction (***Figure 4b***, vertical dotted line indicates the level of prediction and arrows the subsequent updates).

These results indicate that in the network without gating, precision is confounded with prediction and the correlation between precision and effective learning rate is spuriously driven by the prediction itself, whereas in the network with gating, there is a genuine representation of precision beyond the prediction itself, which interacts with the updating of predictions. However, we have so far only provided correlational evidence; to show that the precision represented in the network plays a causal role in the subsequent prediction update, we need to perform an intervention that acts selectively on this precision.

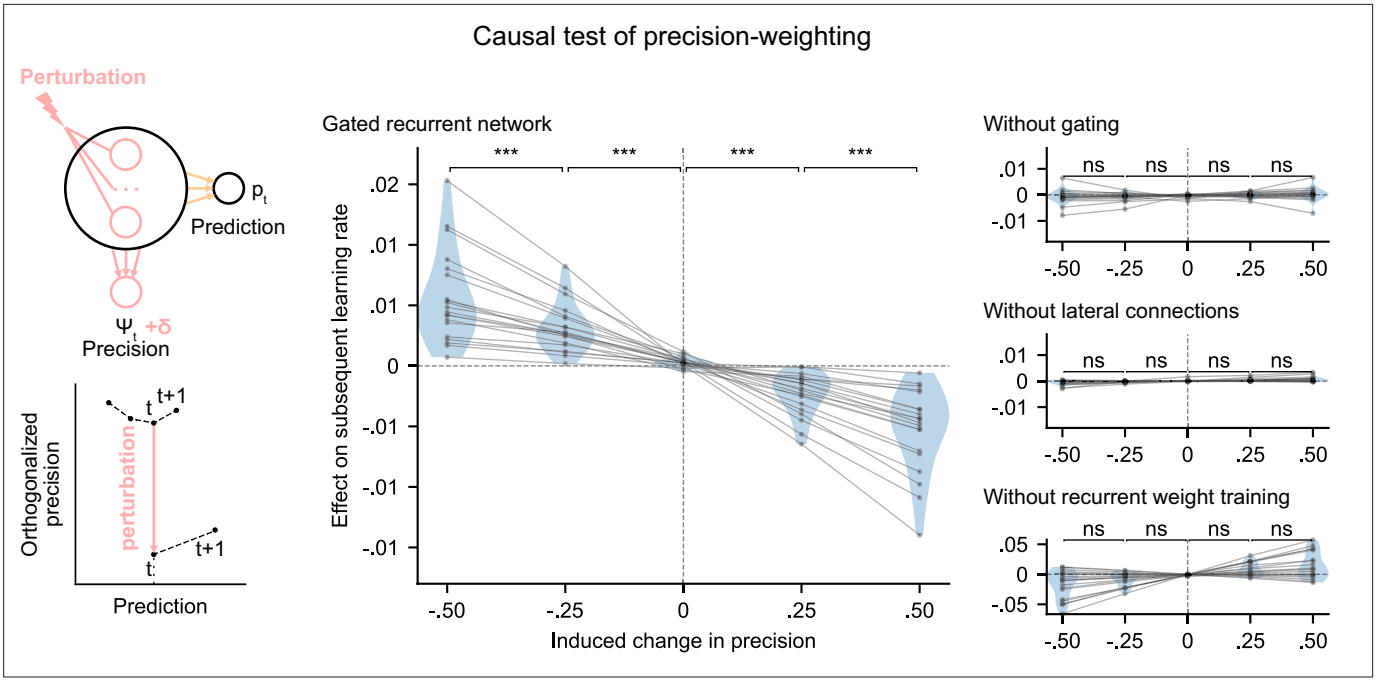


Figure 5. Precision-weighting causally determines the adjustment of the effective learning rate in gated recurrent networks only. Causal test of a network's precision on its effective learning rate. The recurrent activity was perturbed to induce a controlled change δ in the read precision, while keeping the prediction at the current time step—and thus the prediction error at the next time step—constant. This was done by making the perturbation vector orthogonal to the prediction vector and making its projection onto the precision vector equal to δ (bottom left diagram). We measured the perturbation's effect on the subsequent effective learning rate as the difference in learning rate 'with perturbation' minus 'without perturbation' at the next time step (four plots on the right). Each dot (and joining line) corresponds to one network. ***: $p < 0.001$, n.s.: $p > 0.05$ (one-tailed paired t-test).

Causal role of precision-weighting for adaptation to changes

We tested whether the internal representation of precision causally regulated the effective learning rate in the networks using a perturbation experiment. We designed perturbations of the recurrent activity that induced a controlled change in the read precision, while leaving the networks' current prediction unchanged to control for the effect of the prediction error (for the construction of the perturbations, see ***Figure 5*** bottom left diagram and legend, and Materials and methods). These perturbations caused significant changes in the networks' subsequent effective learning rate, commensurate with the induced change in precision, as predicted by the principle of precision-weighting (***Figure 5***, middle plot). Importantly, this causal relationship was abolished in the alternative networks that lacked one of the mechanisms of the gated recurrent architecture (***Figure 5***, right three plots; the slope of the effect was significantly different between the gated recurrent network group and any of the alternative network groups, two-tailed two independent samples t-test, all $t(38) > 4.1$, all $p < 0.001$, all Cohen's $d > 1.3$).

These results show that the gated recurrent networks' ability to adapt to changes indeed relies on their precision-dependent updating and that such precision-weighting does not arise without all three mechanisms of the gated recurrence.

Leveraging and internalizing a latent structure: bigram probabilities

While the changing unigram environment already covers many tasks in the behavioral and neuroscience literature, real-world sequences often exhibit more structure. To study the ability to leverage such structure, we designed a new stochastic and changing environment in which the sequence of observations is no longer generated according to a single unigram probability, p(1), but two 'bigram probabilities' (also known as transition probabilities), p(0|0) and p(1|1), which denote the probability of occurrence of a 0 after a 0 and of a 1 after a 1, respectively (***Figure 6a***; see ***Figure 1—figure supplement 1*** for a graphical model). These bigram probabilities are also changing randomly, with independent change points.

This 'changing bigram environment' is well motivated because there is ample evidence that bigram probabilities play a key role in sequence knowledge in humans and other animals (***Dehaene et al., 2015***) even in the face of changes (***Bornstein and Daw, 2013***; ***Meyniel et al., 2015***).

We assessed how well the networks could leverage the latent bigram structure after having been trained in this environment. For comparison, we tested the optimal agent for this environment as well as two groups of heuristics: delta-rule and leaky estimation of unigram probabilities (as in ***Figure 2b***), and now also delta rule and leaky estimation of bigram probabilities (see Materials and methods for details).

The gated recurrent networks achieved 98% of optimal prediction performance (CI ±0.3%), outperforming the heuristic agents estimating bigram probabilities, and even more so those estimating a unigram probability (***Figure 6c***). To demonstrate that this was due to their internalization of the latent structure, we also tested the gated recurrent networks that had been trained in the changing unigram environment: their performance was much worse (***Figure 6—figure supplement 1***).

At the mechanistic level, all three mechanisms of the gated recurrence are important for this ability to leverage the latent bigram structure. Not only does the performance drop when one of these mechanisms is removed (***Figure 6c***), but also this drop in performance is much larger than that observed in the changing unigram environment (without gating: –11.2% [CI ±1.5% calculated by Welch's t-interval] in the bigram environment vs. –5.5% [CI ±0.6%] in the unigram environment, without lateral connections: –18.5% [CI ±1.8%] vs. –2.9% [CI ±0.2%]; without recurrent weight training: –29.9% [CI ±1.6%] vs. –11.0% [CI ±2.1%]; for every mechanism, there was a significant interaction effect between the removal of the mechanism and the environment on performance, all $F(1,76) > 47.9$, all $p < 0.001$).

Figure 6b illustrates the gated recurrent networks' ability to correctly incorporate the bigram context into its predictions compared to networks lacking one of the mechanisms of the gated recurrence. While a gated recurrent network aptly changes its prediction from one observation to the next according to the preceding observation as the optimal agent does, the other networks fail to show such context-dependent behavior, sometimes even changing their prediction away from the optimal agent.

Altogether these results show that gated recurrent networks can leverage the latent bigram structure, but this ability is impaired when one mechanism of the gated recurrence is missing.

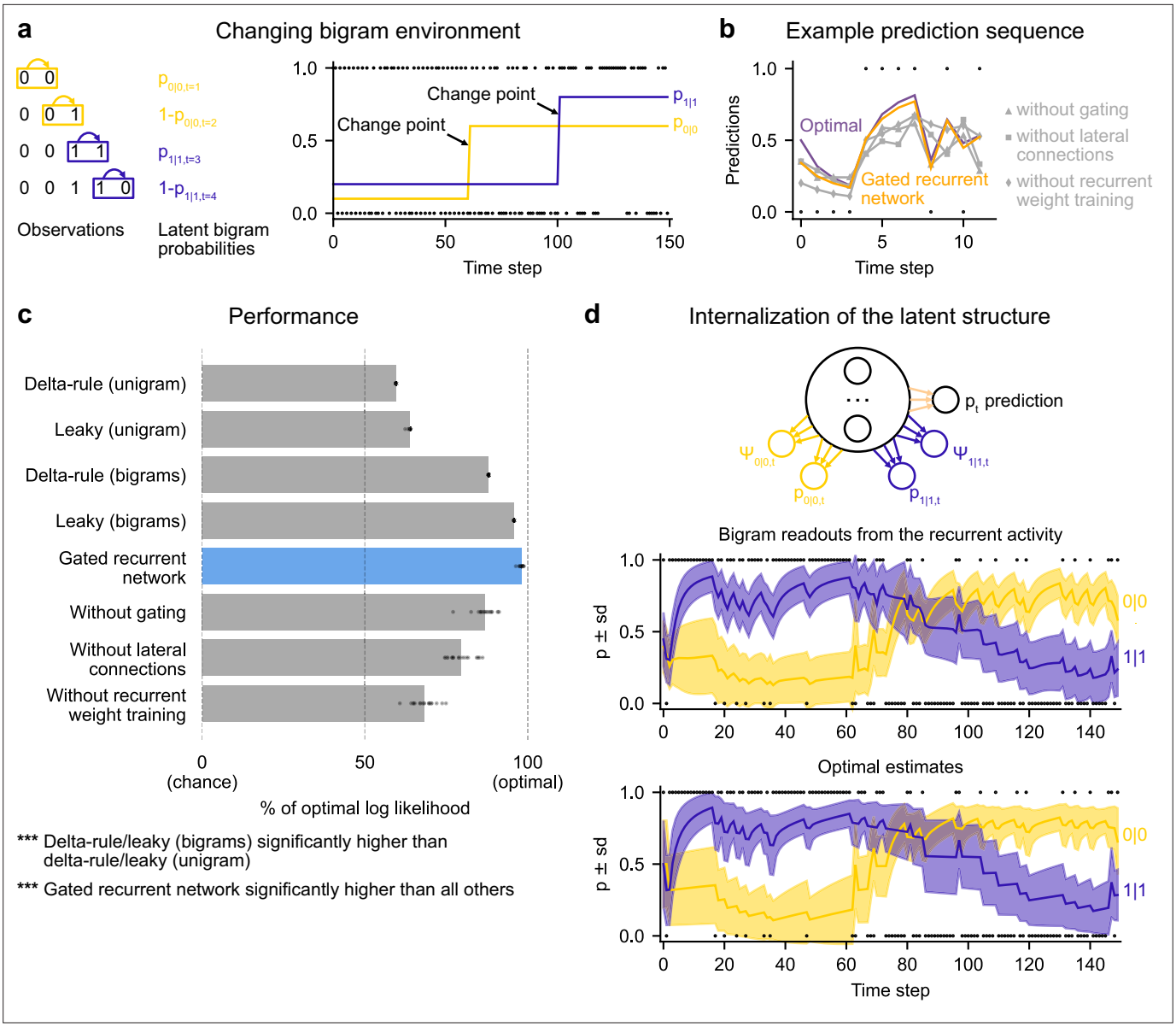


Figure 6. Gated recurrent networks correctly leverage and internalize the latent bigram structure. (**a**) Schematic of the changing bigram environment's latent probabilities (left) and sample generated sequence (right, dots: observations, lines: latent bigram probabilities). At each time step, a binary observation is randomly generated according to the relevant latent bigram probability, $p_{0|0}$ or $p_{1|1}$ depending on the previous observation. $p_{0|0}$ denotes the probability of occurrence of a 0 after a 0 and $p_{1|1}$ that of a 1 after a 1 (note that $p_{1|0}$=1-$p_{0|0}$ and $p_{0|1}$=1-$p_{1|1}$). At any time step, each of the two bigram probabilities can suddenly change to a new value uniformly drawn in [0,1], randomly with a fixed probability and independently from each other. (**b**) Example prediction sequence illustrating each network's ability or inability to change prediction according to the local context, compared to the optimal prediction (dots: observations, lines: predictions). (**c**) Prediction performance of each type of agent in the changing bigram environment. 20 new agents of each type were trained and tested as in ***Figure 2b*** but now in the changing bigram environment (dots: agents; bars: average). The gated recurrent network significantly outperformed every other type of agent ($p < 0.001$, two-tailed two independent samples t-test with Welch's correction for unequal variances). (**d**) Internalization of the latent structure as shown on an out-of-sample sequence: the two bigram probabilities are simultaneously represented in the gated recurrent network (top), and closely follow the optimal estimates (bottom). The readouts were obtained through linear regression from the recurrent activity to four estimates separately: the log odds of the mean and the log precision of the optimal posterior distribution on $p_{0|0}$ and $p_{1|1}$. In (**b**) and (**d**), the networks (out of 20) yielding median performance were selected for illustration purposes.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Performance across training and test environments.

Is the networks' representation of the latent bigram structure impenetrable or easily accessible? We tested the latter possibility by trying to linearly read out the optimal estimate of each of the latent bigram probabilities from the recurrent activity of a gated recurrent network (see Materials and methods). Arguing in favor of an explicit representation, we found that the read estimates of each of the latent bigram probabilities on left-out data were highly accurate (Pearson correlation with the optimal estimates, median and CI: 0.97 [0.97, 0.98] for each of the two bigram probabilities).

In addition to the point estimates of the latent bigram probabilities, we also tested whether a network maintained some information about the precision of each estimate. Again, we assessed the possibility to linearly read out the optimal precision of each estimate and found that the read precisions on left-out data were quite accurate (Pearson correlation with the optimal precisions, median and CI: 0.77 [0.74, 0.78] for one bigram probability and 0.76 [0.74, 0.78] for the other probability).

Figure 6d illustrates the striking resemblance between the estimates read from a gated recurrent network and the optimal estimates. Furthermore, it shows that the network successfully disentangles one bigram probability from the other since the read estimates can evolve independently from each other (for instance during the first 20 time steps, the value for 1|1 changes while the value for 0|0 does not, since only 1s are observed). It is particularly interesting that both bigram probabilities are simultaneously represented, given that only one of them is relevant for the moment-by-moment prediction read by the network's output unit (whose weights cannot change during the sequence).

We conclude that gated recurrent networks internalize the latent bigram structure in such a way that both bigram probabilities are available simultaneously, even though only one of the two is needed at any one time for the prediction.

Leveraging a higher-level structure: inference about latent changes

In real life, latent structures can also exhibit different levels that are organized hierarchically (***Bill et al., 2020***; ***Meyniel et al., 2015***; ***Purcell and Kiani, 2016***). To study the ability to leverage such a hierarchical structure, we designed a third environment in which, in addition to bigram probabilities, we introduced a higher-level factor: the change points of the two bigram probabilities are now coupled, rather than independent as they were in the previous environment (***Figure 7a***; ***Figure 1—figure supplement 1*** shows the hierarchical structure). Due to this coupling, from the agent's point of view, the likelihood that a change point has occurred depends on the observations about both bigrams. Thus, optimal prediction requires the ability to make a higher-level inference: having observed that the frequency of one of the bigrams has changed, one should not only suspect that the latent probability of this bigram has changed but also transfer this suspicion of a change to the latent probability of the other bigram, even without any observations about that bigram.

Such a transfer has been reported in humans (***Heilbron and Meyniel, 2019***, Figure 5B). A typical situation is when a streak of repetitions is encountered (***Figure 7b***): if a long streak of 1 s was deemed unlikely, it should trigger the suspicion of a change point such that p(1|1) is now high, and this suspicion should be transferred to p(0|0) by partially resetting it. This reset is reflected in the change between the prediction following the 0 just before the streak and that following the 0 just after the streak (***Figure 7b***, $|p_{after} - p_{before}|$).

We tested the networks' ability for higher-level inference in the same way, by exposing them to such streaks of repetitions and measuring their change in prediction about the unobserved bigram before and after the streak. More accurately, we compared the change in prediction of the networks trained in the environment with coupled change points to that of the networks trained in the environment with independent change points, since the higher-level inference should only be made in the coupled case.

We found that gated recurrent networks trained in the coupled environment changed their prediction about the unobserved bigram significantly more than networks trained in the independent environment, and this was true across a large range of streak lengths (***Figure 7c***, top plot). The mere presence of this effect is particularly impressive given that the coupling makes very little difference in terms of raw performance (***Figure 6—figure supplement 1***, the networks trained in either the coupled or the independent environment perform very similarly when tested in either environment). All mechanisms of the gated recurrence are important to achieve this higher-level inference since the networks deprived of either gating, lateral connections, or recurrent weight training did not show any effect, no matter the streak length (***Figure 7c***, bottom three plots; for every mechanism, there was a

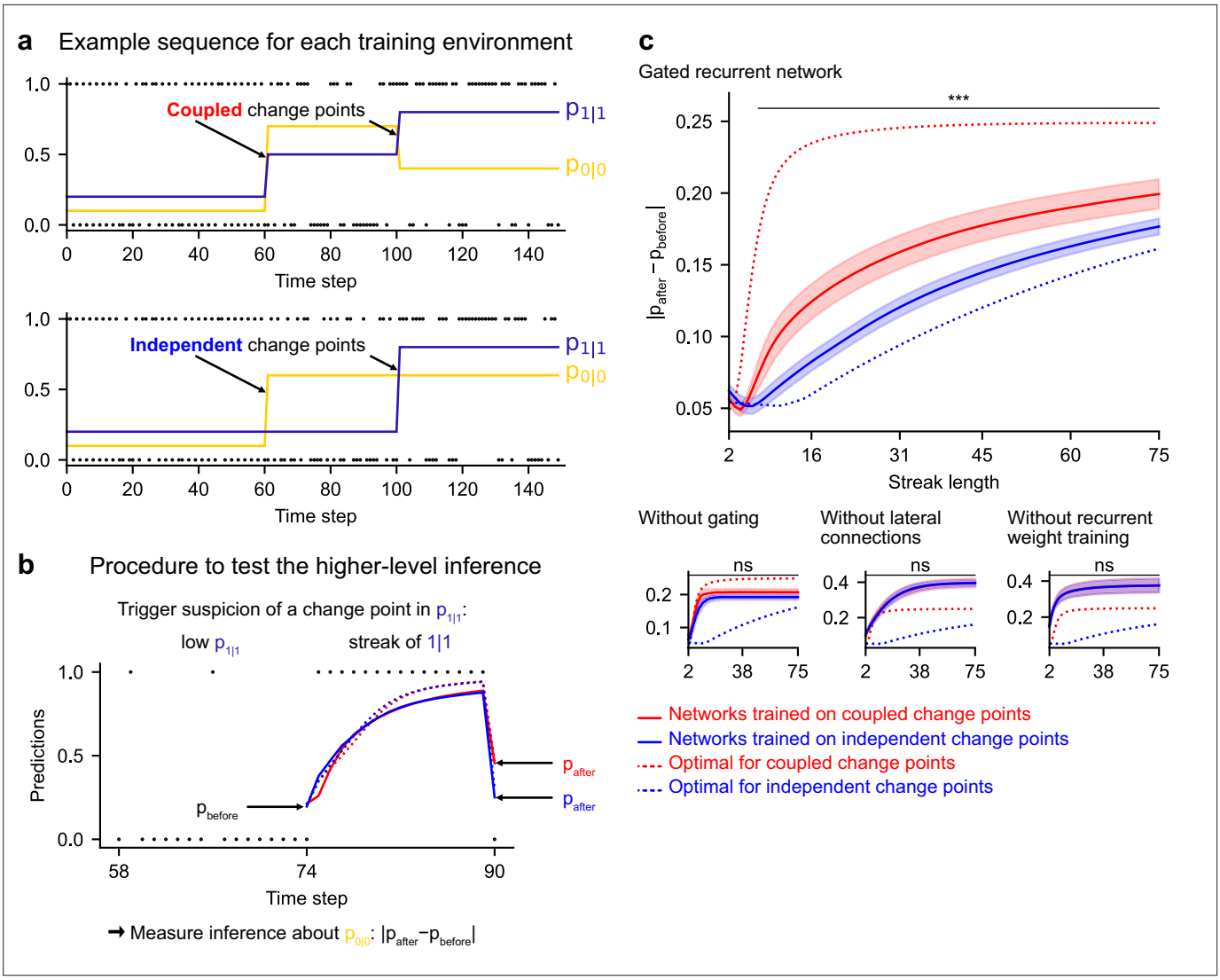


Figure 7. Gated recurrent but not alternative networks leverage a higher-level structure, distinguishing the case where change points are coupled vs. independent. Procedure to test the higher-level inference: (**a**) For each network architecture, 20 networks were trained on sequences where the change points of the two latent bigram probabilities are coupled and 20 other networks were trained on sequences where they are independent (the plots show an example training sequence for each case); (**b**) The networks were then tested on sequences designed to trigger the suspicion of a change point in one bigram probability and measure their inference about the other bigram probability: $|p_{after}-p_{before}|$ should be larger when the agent assumes change points to be coupled rather than independent. The plot shows an example test sequence. Red, blue, solid, and dashed lines: as in (**c**), except that only the gated recurrent network (out of 20) yielding median performance is shown for illustration purposes. (**c**) Change in prediction about the unobserved bigram probability of the networks trained on coupled change points (red) and independent change points (blue) for each network architecture, averaged over sequences. Solid lines and bands show the mean and the 95% confidence interval of the mean over networks. Dotted lines show the corresponding values of the optimal agent for the two cases. Only the gated recurrent architecture yields a significant difference between networks trained on coupled vs. independent change points (one-tailed two independent samples t-test, ***: $p < 0.001$, n.s.: $p > 0.05$).

significant interaction effect between the removal of the mechanism and the training environment on the change in prediction over networks and streak lengths, all $F(1,6076) > 43.2$, all $p < 0.001$).

These results show that gated recurrent networks but not alternative networks leverage the higher level of structure where the change points of the latent probabilities are coupled.

Gated recurrence enables simple solutions

Finally, we highlight the small number of units sufficient to perform quasi-optimally in the increasingly structured environments that we tested: the above-mentioned results were obtained with 11 recurrent units. It turns out that gated recurrent networks can reach a similar performance with even fewer units,

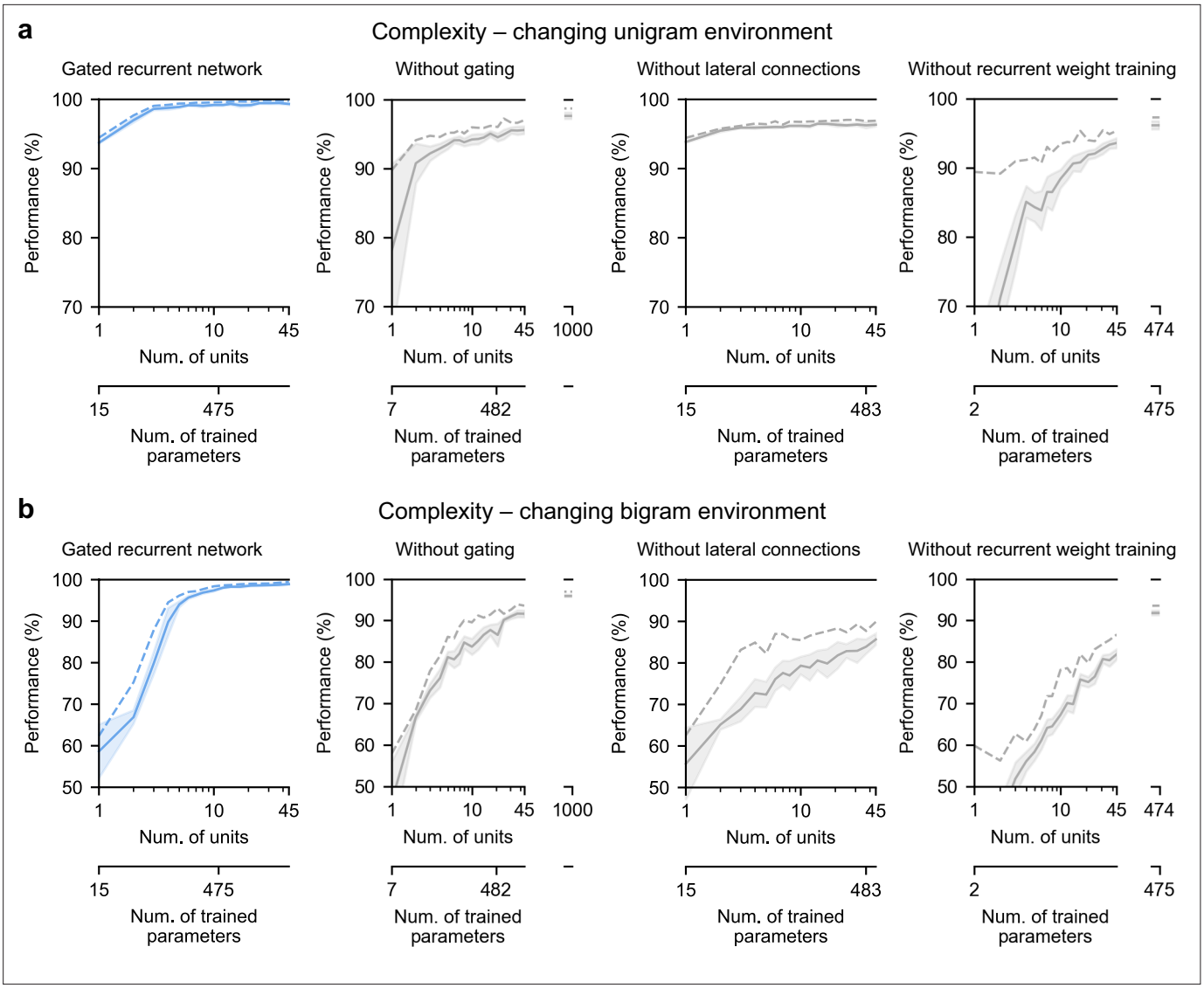


Figure 8. Low-complexity solutions are uniquely enabled by the combination of gating, lateral connections, and recurrent weight training. (**a** and **b**) Prediction performance of each network architecture in the changing unigram environment and the changing bigram environment, respectively, as a function of the number of recurrent units (i.e. space complexity) of the network. For each network architecture and each number of units, 20 networks were trained using hyperparameters that had been optimized prior to training, and prediction performance was measured as the % of optimal log likelihood on new test sequences. Solid lines, bands, and dashed lines show the mean, 95% confidence interval of the mean, and maximum performance, respectively. At the maximum displayed number of units, all of the alternative architectures have exceeded the complexity of the 11-unit gated recurrent network shown on the left and in previous Figures, both in terms of the number of units and the number of trained parameters (indicated on the twin x-axes), but none of them have yet reached its performance.

The online version of this article includes the following figure supplement(s) for figure 8:

Figure supplement 1. Training speed of the gated recurrent networks in the changing unigram and bigram environments.

especially in simpler environments (***Figure 8a and b***, left plot). For instance, in the unigram environment, gated recurrent networks reach 99% of their asymptotic performance with no more than 3 units.

By contrast, without either gating, lateral connections, or recurrent weight training, even when the networks are provided with more units to match the number of trained parameters in the 11-unit gated recurrent networks, they are unable to achieve similar performance (***Figure 8a and b***, right three plots, the twin x-axes indicate the number of units and trained parameters).

With an unlimited number of units, at least in the case without gating (i.e. a vanilla RNN, short for recurrent neural network), the networks will be able to achieve such performance since they are

universal approximators of dynamical systems (***Cybenko, 1989***; ***Schäfer and Zimmermann, 2006***). However, our results indicate that this could require a very large number of units even in the simplest environment tested here (see ***Figure 8a and b***, without gating at 1000 units). Indeed, the slow growth of the vanilla RNNs' performance with the number of units is well described by a power law function, of the form: $(100-p) = c(1/N)^{\alpha}$, where p is the % of optimal performance and N is the number of units. We fitted this law in the unigram environment using the obtained performance from 2 to 45 units and it yielded a goodness-of-fit of $R^2 = 92.4\%$ (fitting was done by linear regression on the logarithm of N and (100−p)). To further confirm the validity of the power law, we then extrapolated to 1,000 units and found that the predicted performance was within 0.2% of the obtained performance for networks of this size (predicted: 97.8%, obtained: 97.6%). Based on this power law, more than 10^4 units would be needed for the vanilla RNN to reach the performance exhibited by the GRU with only 11 units.

Note that, in terms of computational complexity, the number of units is a fair measure of space complexity (i.e. the amount of memory) across the architectures we considered, since in all of them it is equal to the number of state variables (having one state variable per unit, see Materials and methods). What varies across architectures is the number of trained parameters, that is, the degrees of freedom that can be used during training to achieve different dynamics. Still, the conclusion remains the same when an alternative network exceeds the complexity of an 11-unit gated recurrent network in both its number of units and its number of trained parameters.

Therefore, it is the specific computational properties provided by the combination of the three mechanisms that afford effective low-complexity solutions.

Discussion

We have shown that the gated recurrent architecture enables simple and effective solutions: with only 11 units, the networks perform quasi-optimally in environments fraught with randomness, changes, and different levels of latent structure. Moreover, these solutions reproduce several aspects of optimality observed in organisms, including the adaptation of their effective learning rate, the ability to represent the precision of their estimation and to use it to weight their updates, and the ability to represent and leverage the latent structure of the environment. By depriving the architecture of one of its mechanisms, we have shown that three of them are important to achieve such solutions: gating, lateral connections, and the training of recurrent weights.

Can small neural networks behave like Bayesian agents?

A central and much-debated question in the scientific community is whether the brain can perform Bayesian inference (***Knill and Pouget, 2004***; ***Bowers and Davis, 2012***; ***Griffiths et al., 2012***; ***Rahnev and Denison, 2018***; ***Lee and Mumford, 2003***; ***Rao and Ballard, 1999***; ***Sanborn and Chater, 2016***; ***Chater et al., 2006***; ***Findling et al., 2019***; ***Wyart and Koechlin, 2016***; ***Soltani and Izquierdo, 2019***; ***Findling et al., 2021***). From a computational viewpoint, there exists no tractable solution (even approximate) for Bayesian inference in an arbitrary environment, since it is NP-hard (***Cooper, 1990***; ***Dagum and Luby, 1993***). Being a bounded agent (***Simon, 1955***; ***Simon, 1972***), the brain cannot solve Bayesian inference in its most general form. The interesting question is whether the brain can perform Bayesian inference in some environments that occur in real life. More precisely, by 'perform Bayesian inference' one usually means that it performs computations that satisfy certain desirable properties of Bayesian inference, such as taking into account a certain type of uncertainty and a certain type of latent structure (***Courville et al., 2006***; ***Deroy et al., 2016***; ***Griffiths et al., 2012***; ***Knill and Pouget, 2004***; ***Ma, 2010***; ***Ma and Jazayeri, 2014***; ***Tauber et al., 2017***). In this study, we selected specific properties and showed that they can indeed be satisfied when using specific (not all) neural architectures.

In the changing unigram and changing bigram environments, our results provide an existence proof: there exist plausible solutions that are almost indistinguishable from Bayesian inference (i.e. the optimal solution). They exhibit qualitative properties of Bayesian inference that have been demonstrated in humans but are lacking in heuristic solutions, such as the dynamic adjustment of the effective learning rate (***Behrens et al., 2007***; ***Nassar et al., 2010***; ***Nassar et al., 2012***), the internal representation of latent variables and the precision of their estimates (***Boldt et al., 2019***; ***Meyniel et al., 2015***), the precision-weighting of updates (***McGuire et al., 2014***; ***Nassar et al., 2010***; ***Nassar***

et al., 2012), and the ability for higher-level inference (*Bill et al., 2020*; *Heilbron and Meyniel, 2019*; *Purcell and Kiani, 2016*).

The performance we obtained with the gated recurrent architecture is consistent with the numerous other successes it produced in other cognitive neuroscience tasks (*Wang et al., 2018*; *Yang et al., 2019*; *Zhang et al., 2020*). Our detailed study reveals that it offers quasi-optimal low-complexity solutions to new and difficult challenges, including those posed by bigram and higher-level structures and latent probabilities that change unpredictably anywhere in the unit interval. We acknowledge that further generalization to additional challenges remains to be investigated, including the use of more than two categories of observations or continuous observations, and latent structures with longer range dependencies (beyond bigram probabilities).

Minimal set of mechanisms

What are the essential mechanistic elements that enable such solutions? We show that it suffices to have recurrent units of computation equipped with three mechanisms: (1) input, self, and lateral connections which enable each unit to sum up the input with their own and other units' prior value before a non-linear transformation is applied; (2) gating, which enables multiplicative interactions between activities at the summation step; (3) the training of connection weights.

One of the advantages of such mechanisms is their generic character: they do not include any components specifically designed to perform certain probabilistic operations or estimate certain types of latent variables, as often done in neuroscience (*Echeveste et al., 2020*; *Fusi et al., 2007*; *Jazayeri and Movshon, 2006*; *Ma et al., 2006*; *Pecevski et al., 2011*; *Soltani and Wang, 2010*). In addition, they allow adaptive behavior only through recurrent activity dynamics, without involving synaptic plasticity as in other models (*Farashahi et al., 2017*; *Fusi et al., 2005*; *Iigaya, 2016*; *Schultz et al., 1997*). This distinction has implications for the timescale of adaptation: in the brain, recurrent dynamics and synaptic plasticity often involve short and long timescales, respectively. Our study supports this view: recurrent dynamics allow the networks to quickly adapt to a given change in the environment (*Figure 3*), while synaptic plasticity allows the training process to tune the speed of this adaptation to the frequency of change of the environment (*Figure 3—figure supplement 1*).

Our findings suggest that these mechanisms are particularly advantageous to enable solutions with low computational complexity. Without one of them, it seems that a very large number of units (i.e. a large amount of memory) would be needed to achieve comparable performance (*Figure 8*) (note that universal approximation bounds in vanilla RNNs can be very large in terms of number of units [*Barron, 1993*; *Cybenko, 1989*; *Schäfer and Zimmermann, 2006*]). These mechanisms thus seem to be key computational building blocks to build simple and effective solutions. This efficiency can be formalized as the minimum number of units sufficient for near-optimal performance (as in *Orhan and Ma, 2017* who made a similar argument), and it is important for the brain since the brain has limited computational resources (often quantified by the Shannon capacity, i.e. the number of bits that can be transmitted per unit of time, which here amounts to the number of units) (*Bhui et al., 2021*; *Lieder and Griffiths, 2019*). Moreover, simplicity promotes our understanding, and it is with the same goal of understanding that others have used model reduction in large networks (*Dubreuil et al., 2020*; *Jazayeri and Ostojic, 2021*; *Schaeffer et al., 2020*).

Since we cannot exhaustively test all possible parameter values, it might be possible that better solutions exist that were not discovered during training. However, to maximize the chances that the best possible performance is achieved after training, we conducted an extensive hyperparameter optimization, repeated for each environment, architecture, and several number of units, until there is no more improvement according to the Bayesian optimization (see Materials and methods).

Biological implementations of the mechanisms

What biological elements could implement the mechanisms of the gated recurrence? Recurrent connections are ubiquitous in the brain (*Douglas and Martin, 2007*; *Hunt and Hayden, 2017*); the lesser-known aspect is that of gating. In the next paragraph, we speculate on the possible biological implementations of gating, broadly defined as a mechanism that modulates the effective weight of a connection as a function of the network state (and not limited to the very specific form of gating of the GRU).

In neuroscience, many forms of gating have been observed, and they can generally be grouped into three categories according to the neural process that supports them: neural circuits, neural oscillations, and neuromodulation. In neural circuits, a specific pathway can be gated through inhibition/disinhibition by inhibitory (GABAergic) neurons. This has been observed in microscopic circuits, e.g. in pyramidal neurons a dendritic pathway can be gated by interneurons (***Costa et al., 2017***; ***Yang et al., 2016***), or macroscopic circuits, for example in basal ganglia-thalamo-cortical circuits a cortico-cortical pathway can be gated by the basal ganglia and the mediodorsal nucleus of thalamus (***O'Reilly, 2006***; ***O'Reilly and Frank, 2006***; ***Rikhye et al., 2018***; ***Wang and Halassa, 2021***; ***Yamakawa, 2020***). In addition to inhibition/disinhibition, an effective gating can also be achieved by a large population of interacting neurons taking advantage of their nonlinearity (***Beiran et al., 2021***; ***Dubreuil et al., 2020***). Regarding neural oscillations, experiments have shown that activity in certain frequency bands (typically, alpha and beta) can gate behavioral and neuronal responses to the same stimulus (***Baumgarten et al., 2016***; ***Busch et al., 2009***; ***Hipp et al., 2011***; ***Iemi et al., 2019***; ***Klimesch, 1999***; ***Mathewson et al., 2009***). One of the most influential accounts is known as 'pulsed inhibition' (***Hahn et al., 2019***; ***Jensen and Mazaheri, 2010***; ***Klimesch et al., 2007***): a low-frequency signal periodically inhibits a high-frequency signal, effectively silencing the high-frequency signal when the low-frequency signal exceeds a certain threshold. Finally, the binding of certain neuromodulators to the certain receptors of a synapse changes the gain of its input-output transfer function, thus changing its effective weight. This has been demonstrated in neurophysiological studies implicating noradrenaline (***Aston-Jones and Cohen, 2005***; ***Salgado et al., 2016***; ***Servan-Schreiber et al., 1990***), dopamine (***Moyer et al., 2007***; ***Servan-Schreiber et al., 1990***; ***Stalter et al., 2020***; ***Thurley et al., 2008***), and acetylcholine (***Gil et al., 1997***; ***Herrero et al., 2008***) (see review in ***Thiele and Bellgrove, 2018***).

We claim that gated recurrence provides plausible solutions for the brain because its mechanisms can all be biologically implemented and lead to efficient solutions. However, given their multiple biological realizability, the mapping between artificial units and biological neurons is not straightforward: one unit may map to a large population of neurons (e.g. a brain area), or even to a microscopic, subneuronal component (e.g. the dendritic level).

Training: Its role and possible biological counterpart

Regarding the training, our results highlight that it is important to adjust the recurrent weights and thus the network dynamics to the environment (and not fix them as in reservoir computing [***Tanaka et al., 2019***]), but we make no claims about the biological process that leads to such adjustment in brains. It could occur during development (***Sherman et al., 2020***), the life span (***Lillicrap et al., 2020***), or the evolution process (***Zador, 2019***) (these possibilities are not mutually exclusive). Although our training procedure may not be accurate for biology as a whole, two aspects of it may be informative for future research. First, it relies only on the observation sequence (no supervision or reinforcement), leveraging prediction error signals, which have been found in the brain in many studies (***den Ouden et al., 2012***; ***Eshel et al., 2013***; ***Maheu et al., 2019***). Importantly, in predictive coding (***Rao and Ballard, 1999***), the computation of prediction errors is part of the prediction process; here we are suggesting that it may also be part of the training process (as argued in ***O'Reilly et al., 2021***). Second, relatively few iterations of training suffice (***Figure 8—figure supplement 1***, in the order of 10–100; for comparison, ***Wang et al., 2018*** reported training for 40,000 episodes in an environment similar to ours).

Suboptimalities in human behavior

In this study we have focused on some aspects of optimality that humans exhibit in the three environments we explored, but several aspects of their behavior are also suboptimal. In the laboratory, their behavior is often at best qualitatively Bayesian but quantitatively suboptimal. For example, although they adjust their effective learning rate to changes, the base value of their learning rate and their dynamic adjustments may depart from the optimal values (***Nassar et al., 2010***; ***Nassar et al., 2012***; ***Prat-Carrabin et al., 2021***). They may also not update their prediction on every trial, unlike the optimal solution (***Gallistel et al., 2014***; ***Khaw et al., 2017***). Finally, there is substantial interindividual variability which does not exist in the optimal solution (***Khaw et al., 2021***; ***Nassar et al., 2010***; ***Nassar et al., 2012***; ***Prat-Carrabin et al., 2021***). In the future, these suboptimalities could be explored using our networks by making them suboptimal in three ways (among others): by stopping

training before quasi-optimal performance is reached (**Caucheteux and King, 2021**; **Orhan and Ma, 2017**), by constraining the size of the network or its weights (with hard constraints or with regularization penalties) (**Mastrogiuseppe and Ostojic, 2017**; **Sussillo et al., 2015**), or by altering the network in a certain way, such as pruning some of the units or some of the connections (**Blalock et al., 2020**; **Chechik et al., 1999**; **LeCun et al., 1990**; **Srivastava et al., 2014**), or introducing random noise into the activity (**Findling et al., 2021**; **Findling and Wyart, 2020**; **Legenstein and Maass, 2014**). In this way, one could perhaps reproduce the quantitative deviations from optimality while preserving the qualitative aspects of optimality observed in the laboratory.

Implications for experimentalists

If already trained gated recurrent networks exist in the brain, then one can be used in a new but similar enough environment without further training. This is an interesting possibility because, in laboratory experiments mirroring our study, humans perform reasonably well with almost no training but explicit task instructions given in natural language, along with a baggage of prior experience (**Gallistel et al., 2014**; **Heilbron and Meyniel, 2019**; **Khaw et al., 2021**; **Meyniel et al., 2015**; **Peterson and Beach, 1967**). In favor of the possibility to reuse an existing solution, we found that a gated recurrent network can still perform well in conditions different from those it was trained in: across probabilities of change points (**Figure 3—figure supplement 1**) and latent structures (**Figure 6—figure supplement 1**, from bigram to unigram).

In this study, we adopted a self-supervised training paradigm to see if the networks could in principle discover the latent structure from the sequences of observations alone. However, in laboratory experiments, humans often do not have to discover the structure since they are explicitly told what structure they will face and the experiment starts only after ensuring that they have understood it, which makes the comparison to our networks impossible in this setting in terms of training (see similar argument in **Orhan and Ma, 2017**). In the future, it could be interesting to study the ability of gated recurrent networks to switch from one structure to another after having been informed of the current structure as humans do in these experiments. One possible way would be to give a label that indicates the current structure as additional input to our networks, as in **Yang et al., 2019**.

One of our findings may be particularly interesting to experimentalists: in a gated recurrent network, the representations of latent probabilities and the precision of these probability estimates (sometimes referred to as confidence [**Boldt et al., 2019**; **Meyniel et al., 2015**], estimation uncertainty [**McGuire et al., 2014**; **Payzan-LeNestour et al., 2013**], or epistemic uncertainty [**Amini et al., 2020**; **Friston et al., 2015**; **Pezzulo et al., 2015**]) are linearly readable from recurrent activity, the form of decoding most frequently used in neuroscience (**Haxby et al., 2014**; **Kriegeskorte and Diedrichsen, 2019**). These representations arise spontaneously, and their emergence seems to come from the computational properties of gated recurrence together with the need to perform well in a stochastic and changing environment. This yields an empirical prediction: if such networks can be found in the brain, then latent probability estimates and their precision should also be decodable in brain signals, as already found in some studies (**Bach et al., 2011**; **McGuire et al., 2014**; **Meyniel, 2020**; **Meyniel and Dehaene, 2017**; **Payzan-LeNestour et al., 2013**; **Tomov et al., 2020**).

Materials and methods

Sequence prediction problem

The sequence prediction problem to be solved is the following. At each time step, an agent receives as input a binary-valued 'observation', $x_t \in \{0,1\}$, and gives as output a real-valued 'prediction', $p_t \in [0,1]$ which is an estimate of the probability that the value of the next observation is equal to 1, $p(x_{t+1}=1)$. Coding the prediction in terms of the observation being 1 rather than 0 is inconsequential since one can be deduced from the other: $p(x_{t+1}=1) = 1 - p(x_{t+1}=0)$. The agent's objective is to make predictions that maximize the (log) likelihood of observations in the sequence, which technically corresponds to the negative binary cross-entropy cost function:

$$L(p;x) = \sum_{t=0}^{T-1} \log[x_{t+1}p_t + (1-x_{t+1})(1-p_t)] \quad (1)$$

Network architectures

All network architectures consist of a binary input unit, which codes for the current observation, one recurrent layer (sometimes called hidden layer) with a number N of recurrent units, and an output unit, which represents the network's prediction. Unless otherwise stated, N = 11. At every time step, the recurrent unit *i* receives as input the value of the observation, x_t, and the previous activation values of the recurrent units *j* that connect to *i*, $h_{j,t-1}$. It produces as output a new activation value, $h_{i,t}$, which is a real number. The output unit receives as input the activations of all of the recurrent units, and produces as output the prediction p_t.

The parameterized function of the output unit is the same for all network architectures:

$$p_t = \sigma\left(\sum_{i=1}^{N} w_{hp,i} h_{i,t} + b_{hp}\right)$$

where σ is the logistic sigmoid, $w_{hp,i}$ is the weight parameter of the connection from the *i*-th recurrent unit to the output unit, and b_{hp} is the bias parameter of the output unit.

The updating of h_i takes a different form depending on whether gating or lateral connections are included, as described below.

Gated recurrent network

A gated recurrent network includes both gating and lateral connections. This enables multiplicative interactions between the input and recurrent activity as well as the activities of different recurrent units during the updating of h_i. The variant of gating used here is GRU (***Cho et al., 2014***; ***Chung et al., 2014***). For convenience of exposition, we introduce, for each recurrent unit *i*, two intermediate variables in the calculation of the update: the reset gate r_i and the update gate z_i, both of which have their own set of weights and bias. The update gate corresponds to the extent to which a unit can change its values from one time step to the next, and the reset gate corresponds to the balance between recurrent activity and input activity in case of update. Note that r_i and z_i do not count as state variables since the system would be equivalently characterized without them by injecting their expression into the update equation of h_i below. The update is calculated as follows:

$$\begin{aligned}
r_{i,t+1} &= \sigma\left(w_{xr,i} x_{t+1} + b_{xr,i} + w_{hr,ii} h_{i,t} + \textstyle\sum_{j\neq i} w_{hr,ji} h_{j,t} + b_{hr,i}\right)\\
z_{i,t+1} &= \sigma\left(w_{xz,i} x_{t+1} + b_{xz,i} + w_{hz,ii} h_{i,t} + \textstyle\sum_{j\neq i} w_{hz,ji} h_{j,t} + b_{hz,i}\right)\\
h_{i,t+1} &= z_{i,t+1} h_{i,t}\\
&+ (1 - z_{i,t+1})\tanh\left[w_{xh,i} x_{t+1} + b_{xh,i} + r_{i,t+1}(w_{hh,ii} h_{i,t} + \textstyle\sum_{j\neq i} w_{hh,ji} h_{j,t}) + b_{hh,i}\right]\\
h_{i,t=-1} &= 0
\end{aligned}$$

where $(w_{xr,i}, b_{xr,i}, w_{hr,ji}, b_{hr,i})$, $(w_{xz,i}, b_{xz,i}, w_{hz,ji}, b_{hz,i})$, $(w_{xh,i}, b_{xh,i}, w_{hh,ji}, b_{hh,i})$ are the connection weights and biases from the input unit and the recurrent units to unit *i* corresponding to the reset gate, the update gate, and the ungated new activity, respectively.

Another variant of gating is the LSTM (***Hochreiter and Schmidhuber, 1997***). It incorporates similar gating mechanisms as that of the GRU and can achieve the same performance in our task. We chose the GRU because it is simpler than the LSTM and it turned out sufficient.

Without gating

Removing the gating mechanism from the gated recurrent network is equivalent to setting the above variables r_i equal to 1 and z_i equal to 0. This simplifies the calculation of the activations to a single equation, which boils down to a weighted sum of the input and the recurrent units' activity before applying a non-linearity, as follows:

$$h_{i,t+1} = \tanh\left[w_{xh,i} x_{t+1} + b_{xh,i} + w_{hh,ii} h_{i,t} + \sum_{j\neq i} w_{hh,ji} h_{j,t} + b_{hh,i}\right]$$

Another possibility (not considered here) would be to set the value of z_i to a constant other than 1 and treat this value (which amounts to a time constant) as a hyperparameter.

Without lateral connections

Removing lateral connections from the gated recurrent network is equivalent to setting the weights $w_{hr,ji}$, $w_{hz,ji}$, and $w_{hh,ji}$ to 0 for all $j \neq i$. This abolishes the possibility of interaction between recurrent units, which simplifies the calculation of the activations as follows:

$$\begin{aligned} r_{i,t+1} &= \sigma(w_{xr,i}x_{t+1} + b_{xr,i} + w_{hr,ii}h_{i,t} + b_{hr,i}) \\ z_{i,t+1} &= \sigma(w_{xz,i}x_{t+1} + b_{xz,i} + w_{hz,ii}h_{i,t} + b_{hz,i}) \\ h_{i,t+1} &= z_{i,t+1}h_{i,t} + (1 - z_{i,t+1})\tanh[w_{xh,i}x_{t+1} + r_{i,t+1}w_{hh,ii}h_{i,t} + b_{hh,i}] \end{aligned}$$

Note that this architecture still contains gating. We could have tested a simpler architecture without lateral connection and without gating; however, our point is to demonstrate the specific importance of lateral connections to solve the problem we are interested in with few units, and the result is all the more convincing if the network lacking lateral connections has gating (without gating, it would fail even more dramatically).

Without recurrent weight training

The networks referred to as 'without recurrent weight training' have the same architecture as the gated recurrent networks and differ from them only in the way they are trained. While in the other networks, all of the weights and bias parameters are trained, for those networks, only the weights and bias of the output unit, $w_{hp,i}w_{hp,i}$ and b_{hp}, are trained; other weights and biases are fixed to the value drawn at initialization.

Environments

An environment is characterized by its data generating process, that is, the stochastic process used to generate a sequence of observations in that environment. Each of the generative processes is described by a graphical model in ***Figure 1—figure supplement 1*** and further detailed below.

Changing unigram environment

In the changing unigram environment, at each time step, one observation is drawn from a Bernoulli distribution whose probability parameter is the latent variable p_t^{env}. The evolution of this latent variable is described by the following stochastic process.

- Initially, $p_{t=0}^{env}$ is drawn from a uniform distribution on [0,1].
- At the next time step, with probability p_c, p_{t+1}^{env} is drawn anew from a uniform distribution on [0,1] (this event is called a 'change point'), otherwise, p_{t+1}^{env} remains equal to p_t^{env}. The change point probability p_c is fixed in a given environment.

Changing bigram environments

In the changing bigram environments, at each time step, one observation is drawn from a Bernoulli distribution whose probability parameter is either equal to the latent variable $p_{1|1,t}^{env}$, if the previous observation was equal to 1, or to the latent variable $(1 - p_{0|0,t}^{env})$ otherwise (at t = 0, the previous observation is considered to be equal to 0). The evolution of those latent variables is described by a stochastic process which differs depending on whether the change points are independent or coupled.

- In both cases, initially, $p_{0|0,t=0}^{env}$ and $p_{1|1,t=0}^{env}$ are both drawn independently from a uniform distribution on [0,1].
- In the case of *independent change points*, at the next time step, with probability p_c, $p_{0|0,t+1}^{env}$ is drawn anew from a uniform distribution on [0,1], otherwise, $p_{0|0,t+1}^{env}$ remains equal to $p_{0|0,t}^{env}$. Similarly, $p_{1|1,t+1}^{env}$ is either drawn anew with probability p_c or remains equal to $p_{1|1,t}^{env}p_{1|1,t}^{env}$ otherwise, and critically, the occurrence of a change point in $p_{1|1}^{env}$ is independent from the occurrence of a change point in $p_{0|0}^{env}$.
- In the case of *coupled change points*, at the next time step, with probability p_c, $p_{0|0,t+1}^{env}$ and $p_{1|1,t+1}^{env}$ are both drawn anew and independently from a uniform distribution on [0,1], otherwise, both remain equal to $p_{0|0,t}^{env}$ and $p_{1|1,t}^{env}$ respectively.

The changing bigram environment with independent change points and that with coupled change points constitute two distinct environments. When the type of change points is not explicitly mentioned,

the default case is independent change points. For conciseness, we sometimes refer to the changing unigram and changing bigram environments simply as 'unigram' and 'bigram' environments.

In all environments, unless otherwise stated, the length of a sequence is $T = 380$ observations, and the change point probability is $p_c = \frac{1}{75}$, as in previous experiments done with human participants (***Heilbron and Meyniel, 2019***; ***Meyniel et al., 2015***).

Optimal solution

For a given environment among the three possibilities defined above, the optimal solution to the prediction problem can be determined as detailed in ***Heilbron and Meyniel, 2019***. This solution consists in inverting the data-generating process of the environment using Bayesian inference, that is, computing the posterior probability distribution over the values of the latent variables given the history of observation values, and then marginalizing over that distribution to compute the prediction (which is the probability of the next observation given the history of observations). This can be done using a hidden Markov model formulation of the data-generating process where the hidden state includes the values of the latent variables as well as the previous observation in the bigram case, and using the forward algorithm to compute the posterior distribution over the hidden state. Because it would be impossible to compute the probabilities for the infinitely many possible values of the latent variables in the continuous interval [0,1], we discretized the interval into 20 equal-width bins for each of the latent variables. For a more exhaustive treatment, see ***Heilbron and Meyniel, 2019*** and the online code (https://github.com/florentmeyniel/TransitionProbModel).

Heuristic solutions

The four heuristic solutions used here can be classified into 2 × 2 groups depending on:

- which kind of variables are estimated: a *unigram probability* or two *bigram probabilities*.
- which heuristic rule is used in the calculation of the estimates: the *delta-rule* or the *leaky rule*.

The equations used to calculate the estimates are provided below.

Unigram, delta-rule:

$$\widehat{p}_{t+1} = \widehat{p}_t + \alpha(x_{t+1} - \widehat{p}_t)$$

$$\widehat{p}_{t=-1} = 0.5$$

Unigram, leaky rule:

$$n_{0,t+1} = \alpha n_{0,t} + (1 - x_{t+1})$$

$$n_{1,t+1} = \alpha n_{1,t} + x_{t+1}$$

$$n_{0,t=-1} = n_{1,t=-1} = 0$$

$$\widehat{p}_t = \frac{n_{1,t}+1}{n_{1,t}+n_{0,t}+2}$$

Bigrams, delta-rule:

$$\hat{p}_{0|0,t+1} = \hat{p}_{0|0,t} + \alpha(1 - x_t)(1 - x_{t+1} - \hat{p}_{0|0,t})$$

$$\hat{p}_{1|1,t+1} = \hat{p}_{1|1,t} + \alpha x_t(x_{t+1} - \hat{p}_{1|1,t})$$

$$\hat{p}_{0|0,t=-1} = \hat{p}_{1|1,t=-1} = 0.5$$

Bigrams, leaky rule:

$$n_{0|0,t+1} = \alpha n_{0|0,t} + (1 - x_t)(1 - x_{t+1})$$

$$n_{1|0,t+1} = \alpha n_{1|0,t} + (1 - x_t)x_{t+1}$$

$$n_{0|1,t+1} = \alpha n_{0|1,t} + x_t(1 - x_{t+1})$$

$$n_{1|1,t+1} = \alpha n_{1|1,t} + x_t x_{t+1}$$

$$n_{0|0,t=-1} = n_{1|0,t=-1} = n_{0|1,t=-1} = n_{1|1,t=-1} = 0$$

$$\widehat{p}_{0|0,t} = \frac{n_{0|0,t}+1}{n_{0|0,t}+n_{1|0,t}+2}$$

$$\widehat{p}_{1|1,t} = \frac{n_{1|1,t+2}}{n_{1|1,t}+n_{0|1,t}+2}$$

The delta-rule corresponds to the update rule of the Rescorla-Wagner model (***Rescorla and Wagner, 1972***). The leaky rule corresponds to the mean of an approximate posterior which is a Beta distribution whose parameters depend on the leaky counts of observations: $n_1 + 1$ and $n_0 + 1$ (see ***Meyniel et al., 2016*** for more details).

The output prediction value is equal to $\widehat{p}_t$ in the unigram case, and in the bigram case, to $\widehat{p}_{1|1,t}$ if $x_t = 1$ and $(1 - \widehat{p}_{0|0,t})$ otherwise. The parameter α is a free parameter which is trained (using the same training data as the networks) and thus adjusted to the training environment.

Training

For a given environment and a given type of agent among the network types and heuristic types, all the reported results are based on 20 agents, each sharing the same set of hyperparameters and initialized with a different random seed. During training, the parameters of a given agent were adjusted to minimize the binary cross-entropy cost function (see ***Equation 1***). During one iteration of training, the gradients of the cost function with respect to the parameters are computed on a subset of the training data (called a minibatch) using backpropagation through time and are used to update the parameters according to the selected training algorithm. The training algorithm was Adam (***Kingma and Ba, 2015***) for the network types and stochastic gradient descent for the heuristic types.

For the unigram environment, the analyses reported in ***Figures 2–5*** were conducted after training on a common training dataset of 160 minibatches of 20 sequences. For each of the two bigram environments, the analyses reported in ***Figures 6–7*** were conducted after training on a common training dataset (one per environment) of 400 minibatches of 20 sequences. These sizes were sufficient for the validation performance to converge before the end of training for all types of agents.

Table 1. Selected hyperparameter values after optimization.
(*: fixed value.)

Environment	Network architecture	N	η_0	$\sigma_{0,x\cdot}$	$\sigma_{0,h\cdot}$	$\mu_{0,h\cdot,ii}$
unigram	gated recurrent network	3	8.00E-02	0.02	0.02	0*
unigram	gated recurrent network	11	6.60E-02	0.43	0.21	0*
unigram	gated recurrent network	45	4.20E-02	1	0.02	0*
unigram	without gating	3	2.50E-02	1	0.07	0*
unigram	without gating	11	1.70E-02	1	0.07	0*
unigram	without gating	45	7.60E-03	1	0.08	0*
unigram	without gating	1,000	1.34E-04	1	0.04	0*
unigram	without lateral connections	3	5.30E-02	0.02	0.02	1
unigram	without lateral connections	11	2.70E-02	1	0.02	1
unigram	without lateral connections	45	1.30E-02	1	1	1
unigram	without recurrent weight training	3	1.00E-01	1.07	0.55	0*
unigram	without recurrent weight training	11	1.00E-01	2	0.41	0*
unigram	without recurrent weight training	45	1.00E-01	2	0.26	0*
unigram	without recurrent weight training	474	9.60E-03	1	0.1	0*
bigram	gated recurrent network	3	6.30E-02	0.02	1	0*
bigram	gated recurrent network	11	4.40E-02	1	0.02	0*
bigram	gated recurrent network	45	1.60E-02	1	0.02	0*
bigram	without gating	3	5.50E-02	0.02	0.13	0*
bigram	without gating	11	3.20E-02	1	0.05	0*
bigram	without gating	45	8.90E-03	1	0.06	0*
bigram	without gating	1,000	5.97E-05	1	0.03	0*
bigram	without lateral connections	3	4.30E-02	1	0.02	0
bigram	without lateral connections	11	4.30E-02	1	1	0
bigram	without lateral connections	45	2.80E-02	1	1	0
bigram	without recurrent weight training	3	6.60E-02	0.73	0.55	0*
bigram	without recurrent weight training	11	1.00E-01	2	0.45	0*

Parameters initialization

For all of the networks, the bias parameters are randomly initialized from a uniform distribution on $[-1/\sqrt{N}, +1/\sqrt{N}]$ and the weights $w_{hp,i}$ are randomly initialized from a normal distribution with standard deviation $1/\sqrt{N}$ and mean 0. For all the networks, the weights $w_{xr,i}$, $w_{xz,i}$, $w_{xh,i}$ are randomly initialized from a normal distribution with standard deviation $\sigma_{0,x\cdot}$ and mean 0, and the weights $w_{hr,ji}, w_{hz,ji}, w_{hh,ji}$ are randomly initialized from a normal distribution with standard deviation $\sigma_{0,h\cdot}$ and mean 0 for all $j \neq i$ and $\mu_{0,h\cdot,ii}$ for $j = i$. $\sigma_{0,x\cdot}, \sigma_{0,h\cdot}, \mu_{0,h\cdot,ii}$ are hyperparameters that were optimized for a given environment, type of network, and number of units as detailed in the hyperparameter optimization section (the values resulting from this optimization are listed in ***Table 1***).

For the initialization of the parameter α in the heuristic solutions, a random value r is drawn from a log-uniform distribution on the interval $[10^{-2.5}, 10^{-0.5}]$, and the initial value of α is set to r in the delta-rule case or exp(-r) in the leaky rule case.

Hyperparameter optimization

Each type of agent had a specific set of hyperparameters to be optimized. For all network types, it included the initial learning rate of Adam η_0 and the initialization hyperparameters $\sigma_{0,x\cdot}$, $\sigma_{0,h\cdot}$. For the networks without lateral connections specifically, it also included $\mu_{0,h\cdot,ii}$ (for those networks, setting it close to one can help avoid the vanishing gradient problem during training ***Bengio et al., 1994***; ***Sutskever et al., 2013***) for the other networks, this was set to 0. For the heuristic types, it included only the learning rate of the stochastic gradient descent. A unique set of hyperparameter values was determined for each type of agent, each environment, and, for the network types, each number of units, through the optimization described next.

We used Bayesian optimization (***Agnihotri and Batra, 2020***) with Gaussian processes and the upper confidence bound acquisition function to identify the best hyperparameters for each network architecture, environment, and number of units. During the optimization, combinations of hyperparameter values were iteratively sampled, each evaluated over 10 trials with different random seeds, for a total of 60 iterations (hence, 600 trials) for a given architecture, environment, and number of units. In each trial, one network was created, trained, and its cross-entropy was measured on independent test data. The training and test datasets used for the hyperparameter optimization procedure were not used in any other analyses. The training datasets contained respectively 160 and 400 minibatches of 20 sequences for the unigram and the bigram environment; the test datasets contained 200 sequences for each environment. We selected the combination of hyperparameter values corresponding to the iteration that led to the lowest mean test cross-entropy over the 10 trials. The selected values are listed in ***Table 1***.

For the heuristic types, we used random search from a log uniform distribution in the $[10^{-6}, 10^{-1}]$ range over 80 trials to determine the optimal learning rate of the stochastic gradient descent. This led to selecting the value 3.10^{-3} for all heuristic types and all three environments.

Performance analyses

All agents were tested in the environment they were trained in (except for ***Figure 6—figure supplement 1*** which tests cross-environment performance). We used a single test dataset per environment of 1000 sequences independent of the training dataset. The log likelihood L of a given agent was measured from its predictions according to ***Equation 1***. The optimal log likelihood $L_{optimal}$ was measured from the predictions of the optimal solution for the given environment. The chance log likelihood L_{chance} was measured using a constant prediction of 0.5. To facilitate the interpretation of the results, the prediction performance of the agent was expressed as the % of optimal log likelihood, defined as:

$$\frac{L - L_{chance}}{L_{optimal} - L_{chance}} \times 100$$

To test the statistical significance of a comparison of performance between two types of agents, we used a two-tailed two independent samples t-test with Welch's correction for unequal variances.

Analysis of the effective learning rate

The instantaneous effective learning rate of an agent that updates its prediction from p_t to p_{t+1} upon receiving observation x_{t+1} is calculated as:

$$\begin{aligned} \alpha_{t+1} &= \frac{p_{t+1}-p_t}{x_{t+1}-p_t} \\ \alpha_{t=0} &= \frac{p_0-0.5}{x_0-0.5} \end{aligned} \quad (2)$$

We call it 'effective learning rate' because, had the agent been using a delta-rule algorithm, it would be equivalent to the learning rate of the delta-rule (as can be seen by rearranging the above formula into an update equation), and because it can be measured even if the agent uses another algorithm.

Readout analyses

The readout of a given quantity from the recurrent units of a network consists of a weighted sum of the activation values of each unit. To determine the weights of the readout for a given network, we ran a multiple linear regression using, as input variables, the activation of each recurrent unit at a given time step $h_{i,t}$, and as target variable, the desired quantity calculated at the same time step. The regression was run on a training dataset of 900 sequences of 380 observations each (hence, 342,000 samples).

In the unigram environment, the precision readout was obtained using as desired quantity the log precision of the posterior distribution over the unigram variable calculated by the optimal solution as previously described, that is, $\psi_t = -\log \sigma_t$, where σ_t is the standard deviation of the posterior distribution over p_{t+1}^{env}:

$$\sigma_t = \mathbb{SD}[p_{t+1}^{env} | x_0, ..., x_t] \quad (3)$$

In the bigram environment, the readout of the estimate of a given bigram variable was obtained using as desired quantity the log odds of the mean of the posterior distribution over that bigram variable calculated by the optimal solution, and the readout of the precision of that estimate was obtained using the log precision of that same posterior under the above definition of precision.

In ***Figure 4a***, to measure the accuracy of the readout from a given network, we calculated the Pearson correlation between the quantity read from the network and the optimal quantity on a test dataset of 100 sequences (hence, 38,000 samples), independent from any training dataset. To measure the Pearson correlation between the read precision and the subsequent effective learning rate, we used 300 out-of-sample sequences (hence, 114,000 samples). To measure the mutual information between the read precision and the prediction of the network, we also used 300 out-of-sample sequences (114,000 samples).

In ***Figure 6d***, the log odds and log precision were transformed back into mean and standard deviation for visualization purposes.

Dynamics of network activity in the prediction-precision subspace

In ***Figure 4b***, the network activity (i.e. the population activity of the recurrent units in the network) was projected onto the two-dimensional subspace spanned by the prediction vector and the precision vector. The prediction vector is the vector of the weights from the recurrent units to the output unit of the network, w_{hp}. The precision vector is the vector of the weights of the precision readout described above, $w_{h\psi}$. For the visualization, we orthogonalized the precision vector against the prediction vector using the Gram-Schmidt process (i.e. by subtracting from the precision vector its projection onto the prediction vector), and used the orthogonalized precision vector to define the y-axis shown in ***Figure 4b***.

Perturbation experiment to test precision-weighting

The perturbation experiment reported in ***Figure 5*** is designed to test the causal role of the precision read from a given network on its weighting of the next observation, measured through its effective learning rate. We performed this perturbation experiment on each of the 20 networks that were trained within each of the four architectures we considered. The causal instrument is a perturbation vector q that is added to the network's recurrent unit activations. The perturbation vector was randomly generated subject to the following constraints:

- $q \cdot w_{h\psi} = \delta\psi$is the desired change in precision (we used five levels) that is read from the units' activities; it is computed by projecting the perturbation onto the weight vector of the precision readout ($w_{h\psi}$, $\cdot$ is the dot product);

- the perturbation q induces no change in the prediction of the network: $q \cdot w_{hp} = 0$, where w_{hp} is the weight vector of the output unit of the network;
- the perturbation has a constant intensity c across simulations, which we formalize as the norm of the perturbation: $\|q\| = c$.

We describe below the algorithm that we used to generate random perturbations q that satisfy these constraints. The idea is to decompose q into two components: both components leave the prediction unaffected, the first (q_ψ) is used to induce a controlled change in precision, the second (q_r) does not change the precision but is added to ensure a constant intensity of the perturbation across simulations.

1. To ensure no change in precision, we compute Q, the subspace of the activation space spanned by all vectors q that are orthogonal to the prediction weight vector w_{hp}, as the null space of w_{hp} (i.e. the orthogonal complement of the subspace spanned by w_{hp}, dimension N-1).
2. We compute q_ψ, the vector component of Q that affects precision, as the orthogonal projection of $w_{h\psi}$ onto Q (q_ψ is thus collinear to the orthogonalized precision axis shown in ***Figure 4b*** and described above).
3. We compute β_ψ, the coefficient to assign to q_ψ in the perturbation vector to produce the desired change in precision $\delta\psi$, as $\beta_\psi = \frac{\delta\psi}{\|q_\psi \cdot w_{h\psi}\|}$.
4. We compute R, the subspace spanned by all vector components of Q that do not affect precision, as the null space of q_ψ (dimension N-2). A perturbation vector in R therefore leaves both the prediction and the precision unchanged.
5. We draw a random unit vector q_r within R (by drawing from all N-2 components).
6. We compute β_r, the coefficient to assign to q_r in the perturbation vector so as to ensure that the final perturbation's norm equals c, as $\beta_r = \sqrt{c^2 - \beta_\psi^2 \|q_\psi\|^2}$.
7. We combine q_ψ and q_r into the final perturbation vector as $q = \beta_\psi q_\psi + \beta_r q_r$.

The experiment was run on a set of 1000 sample time points randomly drawn from 300 sequences. First, the unperturbed learning rate was measured by running the network on all of the sequences. Second, for each sample time point, the network was run unperturbed up until that point, a perturbation vector was randomly generated for the desired change of precision and applied to the network at that point, then the perturbed network was run on the next time point and its perturbed learning rate was measured. This was repeated for each level of change in precision. Finally, for a given change in precision, the change in learning rate was calculated as the difference between the perturbed and the unperturbed learning rate.

For statistical analysis, we ran a one-tailed paired t-test to test whether the population's mean change in learning rate was higher at one level of precision change than at the next level of precision change. This was done for each of the four consecutive pairs of levels of change in precision.

Test of higher-level inference about changes

For a given network architecture, higher-level inference about changes was assessed by comparing the population of 20 networks trained in the environment with coupled change points to the population of 20 networks trained in the environment with independent change points.

In ***Figure 7c***, the change in unobserved bigram prediction for a given streak length *m* was computed as follows. First, prior sequences were generated and each network was run on each of the sequences. We generated initial sequences of 74 observations each with a probability of 0.2 for the 'observed' bigram (which will render its repetition surprising) and a probability *p* for the 'unobserved' bigram equal to 0.2 or 0.8 (such probabilities, symmetric and substantially different from the default prior 0.5, should render a change in their inferred value detectable). We crossed all possibilities (0|0 or 1|1 as observed bigram, 0.2 or 0.8 for *p*) and generated 100 sequences for each (hence 400 sequences total). Second, at the end of each of these initial sequences, the prediction for the unobserved bigram, p_{before}, was queried by retrieving the output of the network after giving it as input '0' if the unobserved bigram was 0|0 or '1' otherwise. Third, the network was further presented with *m* repeated observations of the same value: '1' if the observed bigram was 1|1 or '0' otherwise. Finally, after this streak of repetition, the new prediction for the unobserved bigram, p_{after}, was queried (as before) and we measured its change with respect to the previous query, $|p_{after} - p_{before}|$. This procedure was repeated for *m* ranging from 2 and 75.

For statistics, we ran a one-tailed two independent samples t-test to test whether the mean change in unobserved bigram prediction of the population trained on coupled change points was higher than that of the population trained on independent change points.

Complexity analyses

The complexity analysis reported in ***Figure 8*** consisted in measuring, for each network architecture and each environment, the performance of optimally trained networks as a function of the number of units N. For optimal training, hyperparameter optimization was repeated at several values of N, for each type of network and each environment (the resulting values are listed in ***Table 1***). For the complexity analysis, a grid of equally spaced N values in logarithmic space between 1 and 45 was generated, an additional value of 474 was included specifically for the networks without recurrent weight training so as to match their number of trained parameters to that of an 11-unit gated recurrent network, and an additional value of 1,000 was included specifically for the networks without gating to facilitate the extrapolation. For every value on this grid, 20 networks of a given architecture in a given environment were randomly initialized with the set of hyperparameter values that was determined to be optimal for the nearest neighboring N value in logarithmic space. The performance of these networks after training was evaluated using a new couple of training and test datasets per environment, each consisting of 400 minibatches of 20 sequences for training and 1000 sequences for testing.

Statistics

To assess the variability between different agent solutions, we trained 20 agents for each type of agent and each environment. These agents have different random seeds (which changes their parameter initialization and how their training data is shuffled). Throughout the article, we report mean or median over these agents, and individual data points when possible or 95% confidence intervals (abbreviated as "CI") otherwise, as fully described in the text and figure legends. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications (***Masse et al., 2019***; ***Yang et al., 2019***). Data analysis was not performed blind to the conditions of the experiments. No data were excluded from the analyses. All statistical tests were two-tailed unless otherwise noted. The data distribution was assumed to be normal, but this was not formally tested. The specific details of each statistical analysis are reported directly in the text.

Code availability

The code to reproduce exhaustively the analyses of this paper is available at https://github.com/cedricfoucault/networks_for_sequence_prediction and archived on Zenodo with DOI: 10.5281/zenodo.5707498. This code also enables to train new networks equipped with any number of units and generate ***Figures 2–7*** with those networks.

Data availability

This paper presents no experimental data. All synthetic data are available in the code repository at https://github.com/cedricfoucault/networks_for_sequence_prediction and archived on Zenodo with DOI: 10.5281/zenodo.5707498.

Acknowledgements

We thank Yair Lakretz for useful feedback, advice, and discussions throughout the project, Alexandre Pouget for his input when starting this project, and Charles Findling for comments on a previous version of the manuscript.

Additional information

Funding

Funder	Grant reference number	Author
École normale supérieure Paris-Saclay	PhD fellowship "Contrat doctoral spécifique normalien"	Cédric Foucault
Agence Nationale de la Recherche	18-CE37-0010-01 "CONFI LEARN"	Florent Meyniel
H2020 European Research Council	ERC StG 947105 "NEURAL PROB"	Florent Meyniel

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Cédric Foucault, Florent Meyniel, Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Visualization, Writing - original draft, Writing - review and editing

Author ORCIDs

Cédric Foucault http://orcid.org/0000-0002-7247-6927
Florent Meyniel http://orcid.org/0000-0002-6992-678X

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.71801.sa1
Author response https://doi.org/10.7554/eLife.71801.sa2

Additional files

Supplementary files

- Transparent reporting form

Data availability

This paper presents no experimental data. All synthetic data are available in the code repository at https://github.com/cedricfoucault/networks_for_sequence_prediction and archived on Zenodo with https://doi.org/10.5281/zenodo.5707498.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Foucault C	2021	Networks for sequence prediction	https://github.com/cedricfoucault/networks_for_sequence_prediction	Github, prediction
Foucault C	2021	Networks for sequence prediction	http://dx.doi.org/10.5281/zenodo.5707498	Zenodo, 10.5281/zenodo.5707498

References

Agnihotri A, Batra N. 2020. Exploring Bayesian Optimization. *Distill* **5**:e26. DOI: https://doi.org/10.23915/distill.00026

Amini A, Schwarting W, Soleimany A, Rus D. 2020. Deep Evidential Regression. Advances in Neural Information Processing Systems. 14927–14937.

Aston-Jones G, Rajkowski J, Kubiak P. 1997. Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. *Neuroscience* **80**:697–715. DOI: https://doi.org/10.1016/s0306-4522(97)00060-2, PMID: 9276487

Aston-Jones G, Cohen JD. 2005. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience* **28**:403–450. DOI: https://doi.org/10.1146/annurev.neuro.28.061604.135709, PMID: 16022602

Bach DR, Hulme O, Penny WD, Dolan RJ. 2011. The known unknowns: neural representation of second-order uncertainty, and ambiguity. *The Journal of Neuroscience* **31**:4811–4820. DOI: https://doi.org/10.1523/JNEUROSCI.1452-10.2011, PMID: 21451019

Barron AR. 1993. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory* **39**:930–945. DOI: https://doi.org/10.1109/18.256500

Baumgarten TJ, Schnitzler A, Lange J. 2016. Prestimulus Alpha Power Influences Tactile Temporal Perceptual Discrimination and Confidence in Decisions. *Cerebral Cortex* **26**:891–903. DOI: https://doi.org/10.1093/cercor/bhu247, PMID: 25331603

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. 2007. Learning the value of information in an uncertain world. *Nature Neuroscience* **10**:1214–1221. DOI: https://doi.org/10.1038/nn1954, PMID: 17676057

Beiran M, Dubreuil A, Valente A, Mastrogiuseppe F, Ostojic S. 2021. Shaping Dynamics With Multiple Populations in Low-Rank Recurrent Networks. *Neural Computation* **33**:1572–1615. DOI: https://doi.org/10.1162/neco_a_01381, PMID: 34496384

Bengio Y, Simard P, Frasconi P. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* **5**:157–166. DOI: https://doi.org/10.1109/72.279181, PMID: 18267787

Berniker M, Kording K. 2008. Estimating the sources of motor errors for adaptation and generalization. *Nature Neuroscience* **11**:1454–1461. DOI: https://doi.org/10.1038/nn.2229, PMID: 19011624

Bhui R, Lai L, Gershman SJ. 2021. Resource-rational decision making. *Current Opinion in Behavioral Sciences* **41**:15–21. DOI: https://doi.org/10.1016/j.cobeha.2021.02.015

Bill J, Pailian H, Gershman SJ, Drugowitsch J. 2020. Hierarchical structure is employed by humans during visual motion perception. *PNAS* **117**:24581–24589. DOI: https://doi.org/10.1073/pnas.2008961117, PMID: 32938799

Blalock D, Ortiz JJG, Frankle J, Guttag J. 2020. What Is the State of Neural Network Pruning. [arXiv]. http://arxiv.org/abs/2003.03033

Boldt A, Blundell C, De Martino B. 2019. Confidence modulates exploration and exploitation in value-based learning. *Neuroscience of Consciousness* **2019**:niz004. DOI: https://doi.org/10.1093/nc/niz004, PMID: 31086679

Bornstein AM, Daw ND. 2013. Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans. *PLOS Computational Biology* **9**:e1003387. DOI: https://doi.org/10.1371/journal.pcbi.1003387, PMID: 24339770

Bowers JS, Davis CJ. 2012. Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin* **138**:389–414. DOI: https://doi.org/10.1037/a0026450, PMID: 22545686

Busch NA, Dubois J, VanRullen R. 2009. The phase of ongoing EEG oscillations predicts visual perception. *The Journal of Neuroscience* **29**:7869–7876. DOI: https://doi.org/10.1523/JNEUROSCI.0113-09.2009, PMID: 19535598

Caucheteux C, King JR. 2021. Language Processing in Brains and Deep Neural Networks: Computational Convergence and Its Limits. [bioRxiv]. DOI: https://doi.org/10.1101/2020.07.03.186288, PMID: 34420675

Chater N, Tenenbaum JB, Yuille A. 2006. Probabilistic models of cognition: conceptual foundations. *Trends in Cognitive Sciences* **10**:287–291. DOI: https://doi.org/10.1016/j.tics.2006.05.007, PMID: 16807064

Chechik G, Meilijson I, Ruppin E. 1999. Neuronal regulation: A mechanism for synaptic pruning during brain maturation. *Neural Computation* **11**:2061–2080. DOI: https://doi.org/10.1162/089976699300016089, PMID: 10578044

Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 1724–1734. DOI: https://doi.org/10.3115/v1/D14-1179

Chung J, Gulcehre C, Cho K, Bengio Y. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. [ArXiv:1412.3555 [Cs]. http://arxiv.org/abs/1412.3555

Cooper GF. 1990. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence* **42**:393–405. DOI: https://doi.org/10.1016/0004-3702(90)90060-D

Costa R, Assael IA, Shillingford B, de Freitas N, Vogels T. 2017. Cortical microcircuits as gated-recurrent neural networks. Advances in Neural Information Processing Systems. .

Courville AC, Daw ND, Touretzky DS. 2006. Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences* **10**:294–300. DOI: https://doi.org/10.1016/j.tics.2006.05.004, PMID: 16793323

Cybenko G. 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems* **2**:303–314. DOI: https://doi.org/10.1007/BF02551274

Dagum P, Luby M. 1993. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence* **60**:141–153. DOI: https://doi.org/10.1016/0004-3702(93)90036-B

de Lange FP, Heilbron M, Kok P. 2018. How Do Expectations Shape Perception? *Trends in Cognitive Sciences* **22**:764–779. DOI: https://doi.org/10.1016/j.tics.2018.06.002, PMID: 30122170

Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C. 2015. The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron* **88**:2–19. DOI: https://doi.org/10.1016/j.neuron.2015.09.019, PMID: 26447569

den Ouden HEM, Kok P, de Lange FP. 2012. How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology* **3**:548. DOI: https://doi.org/10.3389/fpsyg.2012.00548, PMID: 23248610

Deroy O, Spence C, Noppeney U. 2016. Metacognition in Multisensory Perception. *Trends in Cognitive Sciences* **20**:736–747. DOI: https://doi.org/10.1016/j.tics.2016.08.006, PMID: 27612983
Dolan RJ, Dayan P. 2013. Goals and habits in the brain. *Neuron* **80**:312–325. DOI: https://doi.org/10.1016/j.neuron.2013.09.007, PMID: 24139036
Douglas RJ, Martin KAC. 2007. Recurrent neuronal circuits in the neocortex. *Current Biology* **17**:R496-R500. DOI: https://doi.org/10.1016/j.cub.2007.04.024, PMID: 17610826
Dubreuil A, Valente A, Beiran M, Mastrogiuseppe F, Ostojic S. 2020. Complementary Roles of Dimensionality and Population Structure in Neural Computations. [bioRxiv]. DOI: https://doi.org/10.1101/2020.07.03.185942
Echeveste R, Aitchison L, Hennequin G, Lengyel M. 2020. Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nature Neuroscience* **23**:1138–1149. DOI: https://doi.org/10.1038/s41593-020-0671-1, PMID: 32778794
Elman JL. 1990. Finding Structure in Time. *Cognitive Science* **14**:179–211. DOI: https://doi.org/10.1207/s15516709cog1402_1
Elman JL. 1991. Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning* **7**:195–225. DOI: https://doi.org/10.1007/BF00114844
Eshel N, Tian J, Uchida N. 2013. Opening the black box: dopamine, predictions, and learning. *Trends in Cognitive Sciences* **17**:430–431. DOI: https://doi.org/10.1016/j.tics.2013.06.010, PMID: 23830895
Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR. 2001. Efficiency and ambiguity in an adaptive neural code. *Nature* **412**:787–792. DOI: https://doi.org/10.1038/35090500, PMID: 11518957
Farashahi S, Donahue CH, Khorsand P, Seo H, Lee D, Soltani A. 2017. Metaplasticity as a Neural Substrate for Adaptive Learning and Choice under Uncertainty. *Neuron* **94**:401-414.. DOI: https://doi.org/10.1016/j.neuron.2017.03.044, PMID: 28426971
Findling C, Skvortsova V, Dromnelle R, Palminteri S, Wyart V. 2019. Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nature Neuroscience* **22**:2066–2077. DOI: https://doi.org/10.1038/s41593-019-0518-9, PMID: 31659343
Findling C, Wyart V. 2020. Computation Noise Promotes Cognitive Resilience to Adverse Conditions during Decision-Making. [bioRxiv]. DOI: https://doi.org/10.1101/2020.06.10.145300
Findling C, Chopin N, Koechlin E. 2021. Imprecise neural computations as a source of adaptive behaviour in volatile environments. *Nature Human Behaviour* **5**:99–112. DOI: https://doi.org/10.1038/s41562-020-00971-z, PMID: 33168951
Fiser J, Berkes P, Orbán G, Lengyel M. 2010. Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences* **14**:119–130. DOI: https://doi.org/10.1016/j.tics.2010.01.003, PMID: 20153683
Friston K, Rigoli F, Ognibene D, Mathys C, Fitzgerald T, Pezzulo G. 2015. Active inference and epistemic value. *Cognitive Neuroscience* **6**:187–214. DOI: https://doi.org/10.1080/17588928.2015.1020053, PMID: 25689102
Fusi S, Drew PJ, Abbott LF. 2005. Cascade models of synaptically stored memories. *Neuron* **45**:599–611. DOI: https://doi.org/10.1016/j.neuron.2005.02.001, PMID: 15721245
Fusi S, Asaad WF, Miller EK, Wang XJ. 2007. A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales. *Neuron* **54**:319–333. DOI: https://doi.org/10.1016/j.neuron.2007.03.017, PMID: 17442251
Gallistel CR, Krishan M, Liu Y, Miller R, Latham PE. 2014. The perception of probability. *Psychological Review* **121**:96–123. DOI: https://doi.org/10.1037/a0035232, PMID: 24490790
Gijsen S, Grundei M, Lange RT, Ostwald D, Blankenburg F. 2021. Neural surprise in somatosensory Bayesian learning. *PLOS Computational Biology* **17**:e1008068. DOI: https://doi.org/10.1371/journal.pcbi.1008068, PMID: 33529181
Gil Z, Connors BW, Amitai Y. 1997. Differential regulation of neocortical synapses by neuromodulators and activity. *Neuron* **19**:679–686. DOI: https://doi.org/10.1016/s0896-6273(00)80380-3, PMID: 9331357
Griffiths TL, Chater N, Norris D, Pouget A. 2012. How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012). *Psychological Bulletin* **138**:415–422. DOI: https://doi.org/10.1037/a0026884, PMID: 22545687
Hahn G, Ponce-Alvarez A, Deco G, Aertsen A, Kumar A. 2019. Portraits of communication in neuronal networks. *Nature Reviews. Neuroscience* **20**:117–127. DOI: https://doi.org/10.1038/s41583-018-0094-0, PMID: 30552403
Hauser MD, Newport EL, Aslin RN. 2001. Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* **78**:B53-B64. DOI: https://doi.org/10.1016/s0010-0277(00)00132-3, PMID: 11124355
Haxby JV, Connolly AC, Guntupalli JS. 2014. Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience* **37**:435–456. DOI: https://doi.org/10.1146/annurev-neuro-062012-170325, PMID: 25002277
Heilbron M, Meyniel F. 2019. Confidence resets reveal hierarchical adaptive learning in humans. *PLOS Computational Biology* **15**:e1006972. DOI: https://doi.org/10.1371/journal.pcbi.1006972, PMID: 30964861
Herrero JL, Roberts MJ, Delicato LS, Gieselmann MA, Dayan P, Thiele A. 2008. Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. *Nature* **454**:1110–1114. DOI: https://doi.org/10.1038/nature07141, PMID: 18633352
Hipp JF, Engel AK, Siegel M. 2011. Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron* **69**:387–396. DOI: https://doi.org/10.1016/j.neuron.2010.12.027, PMID: 21262474
Hochreiter S, Schmidhuber J. 1997. Long short-term memory. *Neural Computation* **9**:1735–1780. DOI: https://doi.org/10.1162/neco.1997.9.8.1735, PMID: 9377276

Hunt LT, Hayden BY. 2017. A distributed, hierarchical and recurrent framework for reward-based choice. *Nature Reviews. Neuroscience* **18**:172–182. DOI: https://doi.org/10.1038/nrn.2017.7, PMID: 28209978
Iemi L, Busch NA, Laudini A, Haegens S, Samaha J, Villringer A, Nikulin VV. 2019. Multiple mechanisms link prestimulus neural oscillations to sensory responses. *eLife* **8**:e43620. DOI: https://doi.org/10.7554/eLife.43620, PMID: 31188126
Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HEM, Stephan KE. 2013. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* **80**:519–530. DOI: https://doi.org/10.1016/j.neuron.2013.09.009, PMID: 24139048
Iigaya K. 2016. Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *eLife* **5**:e18073. DOI: https://doi.org/10.7554/eLife.18073, PMID: 27504806
Jazayeri M, Movshon JA. 2006. Optimal representation of sensory information by neural populations. *Nature Neuroscience* **9**:690–696. DOI: https://doi.org/10.1038/nn1691, PMID: 16617339
Jazayeri M, Ostojic S. 2021. Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Current Opinion in Neurobiology* **70**:113–120. DOI: https://doi.org/10.1016/j.conb.2021.08.002, PMID: 34537579
Jensen O, Mazaheri A. 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in Human Neuroscience* **4**:186. DOI: https://doi.org/10.3389/fnhum.2010.00186, PMID: 21119777
Kaliukhovich DA, Vogels R. 2014. Neurons in macaque inferior temporal cortex show no surprise response to deviants in visual oddball sequences. *The Journal of Neuroscience* **34**:12801–12815. DOI: https://doi.org/10.1523/JNEUROSCI.2154-14.2014, PMID: 25232116
Khaw MW, Stevens L, Woodford M. 2017. Discrete adjustment to a changing environment: Experimental evidence. *Journal of Monetary Economics* **91**:88–103. DOI: https://doi.org/10.1016/j.jmoneco.2017.09.001
Khaw MW, Stevens L, Woodford M. 2021. Individual differences in the perception of probability. *PLOS Computational Biology* **17**:e1008871. DOI: https://doi.org/10.1371/journal.pcbi.1008871, PMID: 33793574
Kingma DP, Ba J. 2015. Adam: A Method for Stochastic Optimization. 3rd International Conference on Learning Representations, ICLR 2015. .
Klimesch W. 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research. Brain Research Reviews* **29**:169–195. DOI: https://doi.org/10.1016/s0165-0173(98)00056-3, PMID: 10209231
Klimesch W, Sauseng P, Hanslmayr S. 2007. EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Research Reviews* **53**:63–88. DOI: https://doi.org/10.1016/j.brainresrev.2006.06.003, PMID: 16887192
Knill DC, Pouget A. 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences* **27**:712–719. DOI: https://doi.org/10.1016/j.tins.2004.10.007, PMID: 15541511
Kriegeskorte N, Diedrichsen J. 2019. Peeling the Onion of Brain Representations. *Annual Review of Neuroscience* **42**:407–432. DOI: https://doi.org/10.1146/annurev-neuro-080317-061906, PMID: 31283895
LeCun Y, Denker J, Solla S. 1990. Optimal Brain Damage. Advances in Neural Information Processing Systems. .
LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* **521**:436–444. DOI: https://doi.org/10.1038/nature14539, PMID: 26017442
LeCun Y. 2016. Predictive learning. Proc. Speech NIPS. .
Lee TS, Mumford D. 2003. Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision* **20**:1434–1448. DOI: https://doi.org/10.1364/josaa.20.001434, PMID: 12868647
Legenstein R, Maass W. 2014. Ensembles of spiking neurons with noise support optimal probabilistic inference in a dynamically changing environment. *PLOS Computational Biology* **10**:e1003859. DOI: https://doi.org/10.1371/journal.pcbi.1003859, PMID: 25340749
Lieder F, Griffiths TL. 2019. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *The Behavioral and Brain Sciences* **43**:e1. DOI: https://doi.org/10.1017/S0140525X1900061X, PMID: 30714890
Lillicrap TP, Santoro A, Marris L, Akerman CJ, Hinton G. 2020. Backpropagation and the brain. *Nature Reviews. Neuroscience* **21**:335–346. DOI: https://doi.org/10.1038/s41583-020-0277-3, PMID: 32303713
Ma WJ, Beck JM, Latham PE, Pouget A. 2006. Bayesian inference with probabilistic population codes. *Nature Neuroscience* **9**:1432–1438. DOI: https://doi.org/10.1038/nn1790, PMID: 17057707
Ma WJ. 2010. Signal detection theory, uncertainty, and Poisson-like population codes. *Vision Research* **50**:2308–2319. DOI: https://doi.org/10.1016/j.visres.2010.08.035, PMID: 20828581
Ma WJ, Jazayeri M. 2014. Neural coding of uncertainty and probability. *Annual Review of Neuroscience* **37**:205–220. DOI: https://doi.org/10.1146/annurev-neuro-071013-014017, PMID: 25032495
Maheu M, Dehaene S, Meyniel F. 2019. Brain signatures of a multiscale process of sequence learning in humans. *eLife* **8**:e41541. DOI: https://doi.org/10.7554/eLife.41541, PMID: 30714904
Mante V, Sussillo D, Shenoy KV, Newsome WT. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**:78–84. DOI: https://doi.org/10.1038/nature12742, PMID: 24201281
Masse NY, Yang GR, Song HF, Wang XJ, Freedman DJ. 2019. Circuit mechanisms for the maintenance and manipulation of information in working memory. *Nature Neuroscience* **22**:1159–1167. DOI: https://doi.org/10.1038/s41593-019-0414-3, PMID: 31182866
Mastrogiuseppe F, Ostojic S. 2017. Intrinsically-generated fluctuating activity in excitatory-inhibitory networks. *PLOS Computational Biology* **13**:e1005498. DOI: https://doi.org/10.1371/journal.pcbi.1005498, PMID: 28437436

Mathewson KE, Gratton G, Fabiani M, Beck DM, Ro T. 2009. To see or not to see: prestimulus alpha phase predicts visual awareness. *The Journal of Neuroscience* **29**:2725–2732. DOI: https://doi.org/10.1523/JNEUROSCI.3963-08.2009, PMID: 19261866

McGuire JT, Nassar MR, Gold JI, Kable JW. 2014. Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* **84**:870–881. DOI: https://doi.org/10.1016/j.neuron.2014.10.013, PMID: 25459409

Meyniel F, Schlunegger D, Dehaene S. 2015. The Sense of Confidence during Probabilistic Learning: A Normative Account. *PLOS Computational Biology* **11**:e1004305. DOI: https://doi.org/10.1371/journal.pcbi.1004305, PMID: 26076466

Meyniel F, Maheu M, Dehaene S. 2016. Human Inferences about Sequences: A Minimal Transition Probability Model. *PLOS Computational Biology* **12**:e1005260. DOI: https://doi.org/10.1371/journal.pcbi.1005260, PMID: 28030543

Meyniel F, Dehaene S. 2017. Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *PNAS* **114**:E3859–E3868. DOI: https://doi.org/10.1073/pnas.1615773114, PMID: 28439014

Meyniel F. 2020. Brain dynamics for confidence-weighted learning. *PLOS Computational Biology* **16**:e1007935. DOI: https://doi.org/10.1371/journal.pcbi.1007935, PMID: 32484806

Moyer JT, Wolf JA, Finkel LH. 2007. Effects of dopaminergic modulation on the integrative properties of the ventral striatal medium spiny neuron. *Journal of Neurophysiology* **98**:3731–3748. DOI: https://doi.org/10.1152/jn.00335.2007, PMID: 17913980

Nassar MR, Wilson RC, Heasly B, Gold JI. 2010. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience* **30**:12366–12378. DOI: https://doi.org/10.1523/JNEUROSCI.0822-10.2010, PMID: 20844132

Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI. 2012. Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience* **15**:1040–1046. DOI: https://doi.org/10.1038/nn.3130, PMID: 22660479

Orhan AE, Ma WJ. 2017. Efficient probabilistic inference in generic neural networks trained with non-probabilistic feedback. *Nature Communications* **8**:1–14. DOI: https://doi.org/10.1038/s41467-017-00181-8, PMID: 28743932

O'Reilly RC. 2006. Biologically based computational models of high-level cognition. *Science* **314**:91–94. DOI: https://doi.org/10.1126/science.1127242, PMID: 17023651

O'Reilly RC, Frank MJ. 2006. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation* **18**:283–328. DOI: https://doi.org/10.1162/089976606775093909, PMID: 16378516

O'Reilly RC, Russin JL, Zolfaghar M, Rohrlich J. 2021. Deep Predictive Learning in Neocortex and Pulvinar. *Journal of Cognitive Neuroscience* **33**:1158–1196. DOI: https://doi.org/10.1162/jocn_a_01708, PMID: 34428793

Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty JP. 2013. The neural representation of unexpected uncertainty during value-based decision making. *Neuron* **79**:191–201. DOI: https://doi.org/10.1016/j.neuron.2013.04.037, PMID: 23849203

Pecevski D, Buesing L, Maass W. 2011. Probabilistic inference in general graphical models through sampling in stochastic networks of spiking neurons. *PLOS Computational Biology* **7**:e1002294. DOI: https://doi.org/10.1371/journal.pcbi.1002294, PMID: 22219717

Peterson CR, Beach LR. 1967. Man as an intuitive statistician. *Psychological Bulletin* **68**:29–46. DOI: https://doi.org/10.1037/h0024722, PMID: 6046307

Pezzulo G, Rigoli F, Friston K. 2015. Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology* **134**:17–35. DOI: https://doi.org/10.1016/j.pneurobio.2015.09.001, PMID: 26365173

Prat-Carrabin A, Wilson RC, Cohen JD, Azeredo da Silveira R. 2021. Human inference in changing environments with temporal structure. *Psychological Review* **128**:879–912. DOI: https://doi.org/10.1037/rev0000276, PMID: 34516148

Purcell BA, Kiani R. 2016. Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. *PNAS* **113**:E4531-E4540. DOI: https://doi.org/10.1073/pnas.1524685113, PMID: 27432960

Rahnev D, Denison RN. 2018. Suboptimality in perceptual decision making. *The Behavioral and Brain Sciences* **41**:e223. DOI: https://doi.org/10.1017/S0140525X18000936, PMID: 29485020

Rao RP, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* **2**:79–87. DOI: https://doi.org/10.1038/4580, PMID: 10195184

Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical Conditioning II: Current Research and Theory. 64–99.

Rikhye RV, Gilra A, Halassa MM. 2018. Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nature Neuroscience* **21**:1753–1763. DOI: https://doi.org/10.1038/s41593-018-0269-z, PMID: 30455456

Robinson JG. 1979. An analysis of the organization of vocal communication in the titi monkey Callicebus moloch. *Zeitschrift Fur Tierpsychologie* **49**:381–405. DOI: https://doi.org/10.1111/j.1439-0310.1979.tb00300.x, PMID: 115173

Rose GJ, Goller F, Gritton HJ, Plamondon SL, Baugh AT, Cooper BG. 2004. Species-typical songs in white-crowned sparrows tutored with only phrase pairs. *Nature* **432**:753–758. DOI: https://doi.org/10.1038/nature02992, PMID: 15592413

Saffran JR, Aslin RN, Newport EL. 1996. Statistical learning by 8-month-old infants. *Science* **274**:1926–1928. DOI: https://doi.org/10.1126/science.274.5294.1926, PMID: 8943209

Sahani M, Dayan P. 2003. Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity. *Neural Computation* **15**:2255–2279. DOI: https://doi.org/10.1162/089976603322362356, PMID: 14511521

Salgado H, Treviño M, Atzori M. 2016. Layer- and area-specific actions of norepinephrine on cortical synaptic transmission. *Brain Research* **1641**:163–176. DOI: https://doi.org/10.1016/j.brainres.2016.01.033, PMID: 26820639

Sanborn AN, Chater N. 2016. Bayesian Brains without Probabilities. *Trends in Cognitive Sciences* **20**:883–893. DOI: https://doi.org/10.1016/j.tics.2016.10.003, PMID: 28327290

Saxe A, Nelli S, Summerfield C. 2021. If deep learning is the answer, what is the question? *Nature Reviews. Neuroscience* **22**:55–67. DOI: https://doi.org/10.1038/s41583-020-00395-8, PMID: 33199854

Schaeffer R, Khona M, Meshulam L, Laboratory IB, Fiete IR. 2020. Reverse-engineering Recurrent Neural Network solutions to a hierarchical inference task for mice. NeurIPS ProceedingsSearch. . DOI: https://doi.org/10.1101/2020.06.09.142745

Schäfer AM, Zimmermann HG. 2006. Recurrent Neural Networks Are Universal Approximators. Kollias SD, Stafylopatis A, Duch W, Oja E (Eds). *Artificial Neural Networks – ICANN 2006*. Berlin Heidelberg: Springer. p. 632–640. DOI: https://doi.org/10.1007/11840817

Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM. 2013. Neural representations of events arise from temporal community structure. *Nature Neuroscience* **16**:486–492. DOI: https://doi.org/10.1038/nn.3331, PMID: 23416451

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. *Science* **275**:1593–1599. DOI: https://doi.org/10.1126/science.275.5306.1593, PMID: 9054347

Servan-Schreiber D, Printz H, Cohen JD. 1990. A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science* **249**:892–895. DOI: https://doi.org/10.1126/science.2392679, PMID: 2392679

Sherman BE, Graves KN, Turk-Browne NB. 2020. The prevalence and importance of statistical learning in human cognition and behavior. *Current Opinion in Behavioral Sciences* **32**:15–20. DOI: https://doi.org/10.1016/j.cobeha.2020.01.015, PMID: 32258249

Simon HA. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics* **69**:99. DOI: https://doi.org/10.2307/1884852

Simon HA. 1972. Theories of bounded rationality. *Decision and Organization* **1**:161–176.

Sohn H, Narain D, Meirhaeghe N, Jazayeri M. 2019. Bayesian Computation through Cortical Latent Dynamics. *Neuron* **103**:934-947.. DOI: https://doi.org/10.1016/j.neuron.2019.06.012, PMID: 31320220

Soltani A, Wang XJ. 2010. Synaptic computation underlying probabilistic inference. *Nature Neuroscience* **13**:112–119. DOI: https://doi.org/10.1038/nn.2450, PMID: 20010823

Soltani A, Izquierdo A. 2019. Adaptive learning under expected and unexpected uncertainty. *Nature Reviews. Neuroscience* **20**:635–644. DOI: https://doi.org/10.1038/s41583-019-0180-y, PMID: 31147631

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. 2014. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* **15**:1929–1958.

Stalter M, Westendorff S, Nieder A. 2020. Dopamine Gates Visual Signals in Monkey Prefrontal Cortex Neurons. *Cell Reports* **30**:164-172.. DOI: https://doi.org/10.1016/j.celrep.2019.11.082, PMID: 31914383

Sterling P. 2004. Principles of allostasis: Optimal design, predictive regulation, pathophysiology, and rational therapeutics. Allostasis, Homeostasis, and the Costs of Physiological Adaptation. 17–64. DOI: https://doi.org/10.1017/CBO9781316257081

Summerfield C, de Lange FP. 2014. Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews. Neuroscience* **15**:745–756. DOI: https://doi.org/10.1038/nrn3838, PMID: 25315388

Sussillo D, Churchland MM, Kaufman MT, Shenoy KV. 2015. A neural network that finds a naturalistic solution for the production of muscle activity. *Nature Neuroscience* **18**:1025–1033. DOI: https://doi.org/10.1038/nn.4042, PMID: 26075643

Sutskever I, Martens J, Dahl G, Hinton G. 2013. On the importance of initialization and momentum in deep learning. International Conference on Machine Learning. 1139–1147.

Sutton R. 1992. Gain Adaptation Beats Least Squares. In Proceedings of the 7th Yale Workshop on Adaptive and Learning Systems. 161–166.

Sutton RS, Barto AG. 1998. Introduction to Reinforcement Learning. MIT Press. DOI: https://doi.org/10.1109/TNN.1998.712192

Tanaka G, Yamane T, Héroux JB, Nakane R, Kanazawa N, Takeda S, Numata H, Nakano D, Hirose A. 2019. Recent advances in physical reservoir computing: A review. *Neural Networks* **115**:100–123. DOI: https://doi.org/10.1016/j.neunet.2019.03.005, PMID: 30981085

Tauber S, Navarro DJ, Perfors A, Steyvers M. 2017. Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review* **124**:410–441. DOI: https://doi.org/10.1037/rev0000052, PMID: 28358549

Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. 2011. How to grow a mind: statistics, structure, and abstraction. *Science* **331**:1279–1285. DOI: https://doi.org/10.1126/science.1192788, PMID: 21393536

Thiele A, Bellgrove MA. 2018. Neuromodulation of Attention. *Neuron* **97**:769–785. DOI: https://doi.org/10.1016/j.neuron.2018.01.008, PMID: 29470969

Thurley K, Senn W, Lüscher HR. 2008. Dopamine increases the gain of the input-output response of rat prefrontal pyramidal neurons. *Journal of Neurophysiology* **99**:2985–2997. DOI: https://doi.org/10.1152/jn.01098.2007, PMID: 18400958

Todd PM, Gigerenzer G. 2000. Précis of Simple heuristics that make us smart. *The Behavioral and Brain Sciences* **23**:727–741; . DOI: https://doi.org/10.1017/s0140525x00003447, PMID: 11301545

Tomov MS, Truong VQ, Hundia RA, Gershman SJ. 2020. Dissociable neural correlates of uncertainty underlie different exploration strategies. *Nature Communications* **11**:2371. DOI: https://doi.org/10.1038/s41467-020-15766-z, PMID: 32398675

Ulanovsky N, Las L, Farkas D, Nelken I. 2004. Multiple time scales of adaptation in auditory cortex neurons. *The Journal of Neuroscience* **24**:10440–10453. DOI: https://doi.org/10.1523/JNEUROSCI.1905-04.2004, PMID: 15548659

Vinckier F, Gaillard R, Palminteri S, Rigoux L, Salvador A, Fornito A, Adapa R, Krebs MO, Pessiglione M, Fletcher PC. 2016. Confidence and psychosis: a neuro-computational account of contingency learning disruption by NMDA blockade. *Molecular Psychiatry* **21**:946–955. DOI: https://doi.org/10.1038/mp.2015.73, PMID: 26055423

Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M. 2018. Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience* **21**:860–868. DOI: https://doi.org/10.1038/s41593-018-0147-8, PMID: 29760527

Wang MB, Halassa MM. 2021. Thalamocortical Contribution to Solving Credit Assignment in Neural Systems. [arXiv]. http://arxiv.org/abs/2104.01474

Wark B, Fairhall A, Rieke F. 2009. Timescales of inference in visual adaptation. *Neuron* **61**:750–761. DOI: https://doi.org/10.1016/j.neuron.2009.01.019, PMID: 19285471

Wolpert DM, Ghahramani Z, Jordan MI. 1995. An internal model for sensorimotor integration. *Science* **269**:1880–1882. DOI: https://doi.org/10.1126/science.7569931, PMID: 7569931

Wyart V, Koechlin E. 2016. Choice variability and suboptimality in uncertain environments. *Current Opinion in Behavioral Sciences* **11**:109–115. DOI: https://doi.org/10.1016/j.cobeha.2016.07.003

Yamakawa H. 2020. Attentional Reinforcement Learning in the Brain. *New Generation Computing* **38**:49–64. DOI: https://doi.org/10.1007/s00354-019-00081-z

Yang GR, Murray JD, Wang XJ. 2016. A dendritic disinhibitory circuit mechanism for pathway-specific gating. *Nature Communications* **7**:12815. DOI: https://doi.org/10.1038/ncomms12815

Yang GR, Joglekar MR, Song HF, Newsome WT, Wang XJ. 2019. Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience* **22**:297–306. DOI: https://doi.org/10.1038/s41593-018-0310-2, PMID: 30643294

Yu AJ, Cohen JD. 2008. Sequential effects: Superstition or rational behavior?. Advances in neural information processing systems. 1873–1880 PMID: 26412953.

Zador AM. 2019. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications* **10**:3770. DOI: https://doi.org/10.1038/s41467-019-11786-6, PMID: 31434893

Zhang Z, Cheng H, Yang T. 2020. A recurrent neural network framework for flexible and adaptive decision making based on sequence learning. *PLOS Computational Biology* **16**:e1008342. DOI: https://doi.org/10.1371/journal.pcbi.1008342, PMID: 33141824

Figures and figure supplements

Gated recurrence enables simple and accurate sequence prediction in stochastic, changing, and structured environments

Cédric Foucault and Florent Meyniel

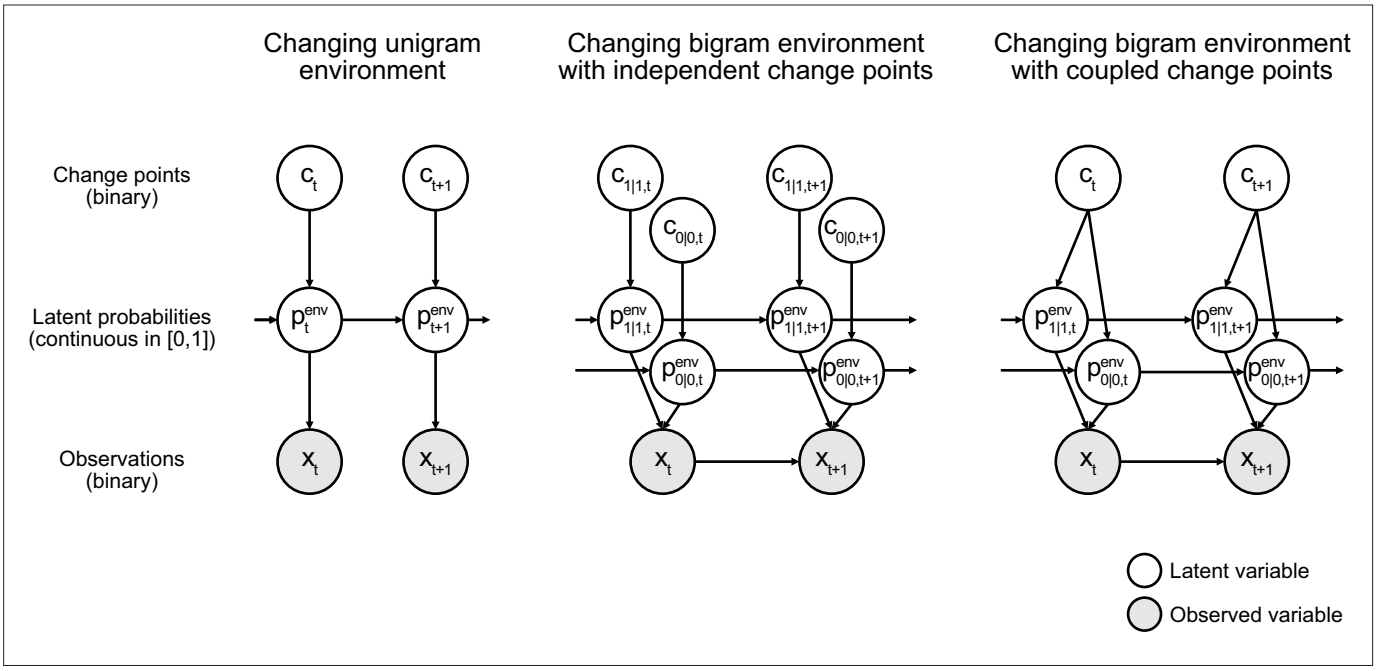


Figure 1—figure supplement 1. Graphical model of the generative process of each environment. Nodes encode the variables and edges the conditional dependencies between variables. Each graph represents a factorization of the joint probability distribution of all variables in the generative process: this joint distribution is the product of the conditional probability distributions of each variable given its parents in the graph. For further details on the generative processes, see Materials and methods. In all environments, inferring the next observation from previous observations using such a graph is computationally difficult because it requires computing and marginalizing over the continuous probability distribution of the latent probabilities. This distribution is not easy to compute because it incorporates the likelihoods of the observations (for any latent probability value) and the change point probabilities from all previous time steps, and requires normalization. Notice also the increasingly complex conditional structures of the graphs from left to right. In the unigram environment, observations are conditionally independent given the latent probabilities, but in the bigram environments, they interact. In the bigram environment with coupled change points, the hierarchical structure implies that the two latent bigram probabilities are no longer conditionally independent of each other given their values at the previous time step, since they are connected by a common parent (the change point).

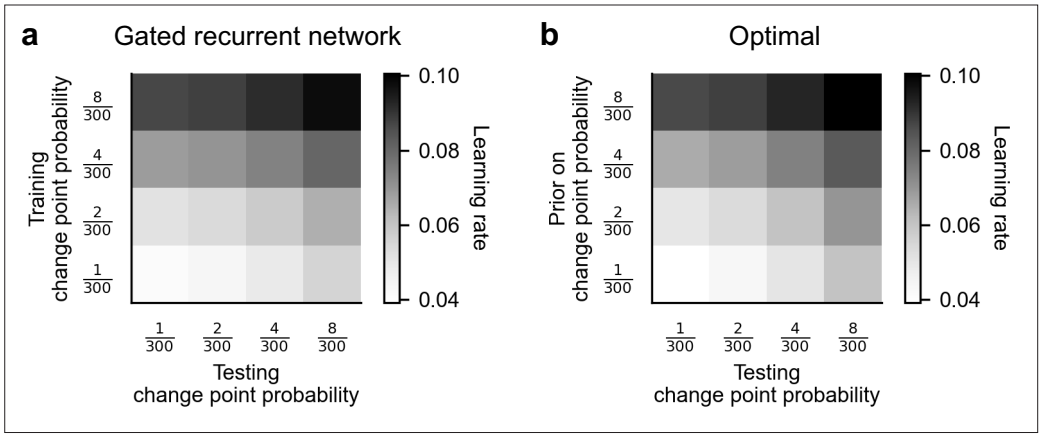


Figure 3—figure supplement 1. Attunement of the effective learning rate to the change point probabilities. (**a**) Average effective learning rate of the gated recurrent networks as a function of the change point probability used during testing (columns) and during training (rows). Each row corresponds to a different set of 20 networks trained in the changing unigram environment with the indicated change point probability. Each column corresponds to a different test set with the indicated change point probability, each of 1000 out-of-sample sequences. The networks' effective learning rate was measured and averaged over time, sequences, and networks. (**b**) Average effective learning rate of the optimal agent as a function of the change point probability used during testing (columns) and the prior on the change point probability assumed by the model (rows). The optimal agent was tested on the same sets of sequences as the gated recurrent networks and its effective learning rate was averaged over time and sequences.

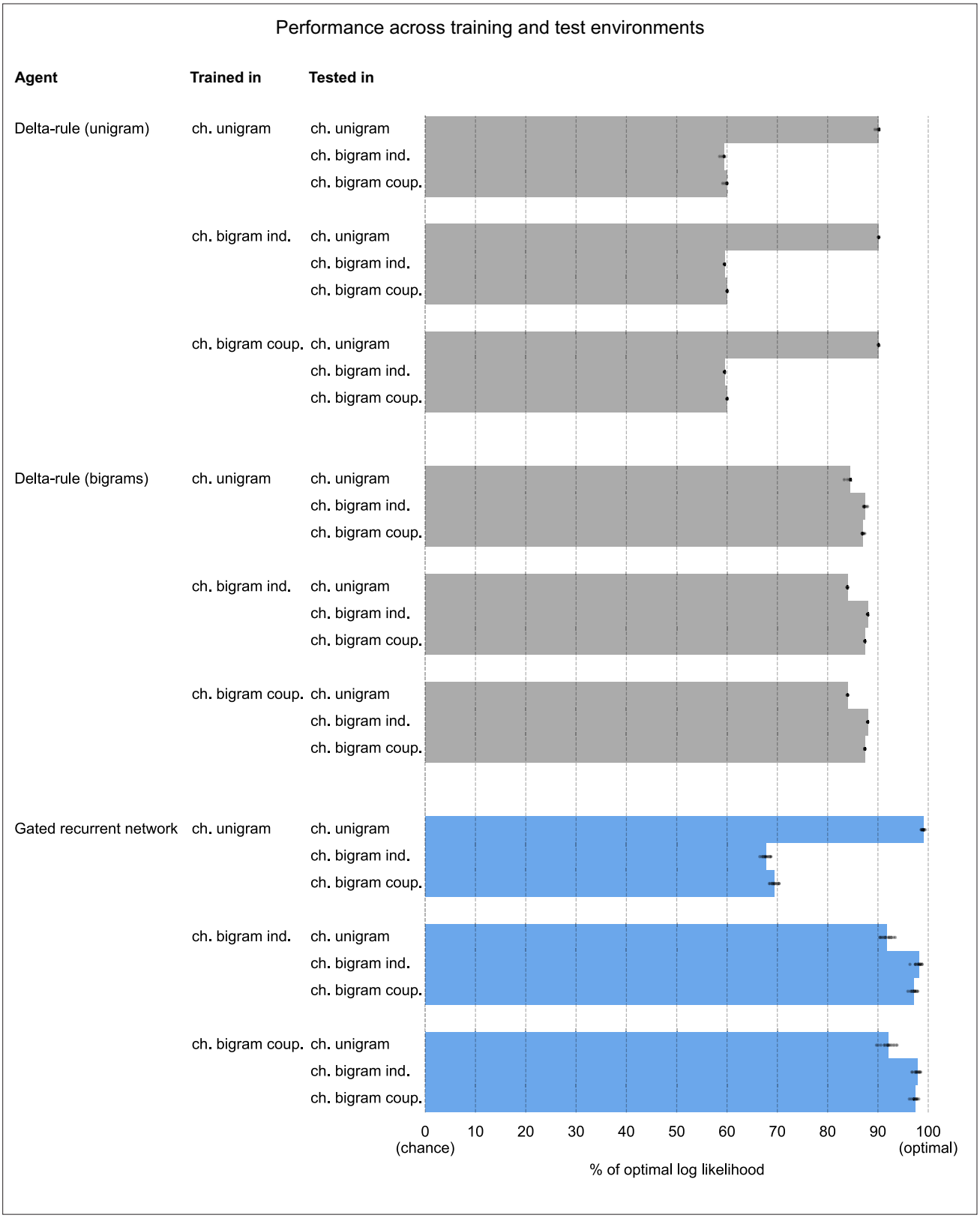


Figure 6—figure supplement 1. Performance across training and test environments. For each type of agent and each environment, a set of 20 agents was trained in the given environment as in ***Figures 2, 5 and 6***. The performance of each set of trained agents was then evaluated in each test environment, using 1,000 new sequences per environment and the same performance measure as in ***Figures 2 and 5***. ch.: changing; ind.: independent change points; coup: coupled change points.

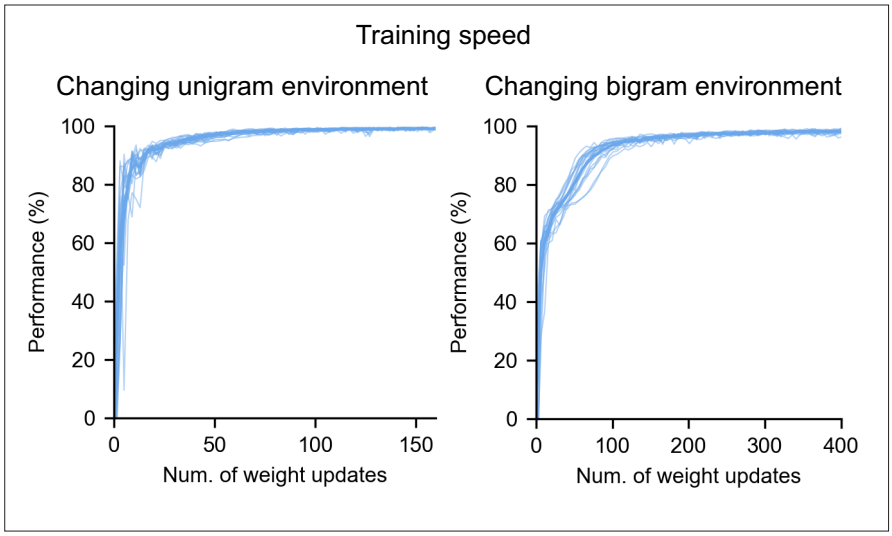


Figure 8—figure supplement 1. Training speed of the gated recurrent networks in the changing unigram and bigram environments. During training, the networks' weights were iteratively updated, with each update based on the evaluation of the cost function on 20 sequences. Prediction performance was repeatedly measured after each iteration as the % of optimal log likelihood on an out-of-sample validation set of 200 sequences. The thin lines and the thick line show the mean and the individual performances of the 20 networks, respectively.

Chapter IV: Article 3, fMRI study (Foucault*, Bounmy*, et al., in preparation)

Preprint: https://www.biorxiv.org/content/10.1101/2024.02.28.582455

A nonlinear code for event probability in the human brain

Cedric Foucault[1,2,*,§], Tiffany Bounmy[1,*], Sébastien Demortain[1], Bertrand Thiririon[3], Evelyn Eger[1], Florent Meyniel[1,4,§]

1. Cognitive Neuroimaging Unit, NeuroSpin (INSERM-CEA), University of Paris-Saclay, 91191 Gif-sur-Yvette, France
2. Sorbonne University, Doctoral College, F-75005 Paris, France
3. Inria, CEA, University of Paris-Saclay, Palaiseau, France
4. GHU Paris, psychiatrie et neurosciences, Hôpital Sainte-Anne, Institut de neuromodulation, 75014 Paris, France

*: co-first authors, §: corresponding authors cedric.foucault@gmail.com and florent.meyniel@cea.fr.

This version is a working version, it has not been submitted to a journal.

Abstract

Assessing probabilities and predicting future events are fundamental for perception and adaptive behavior, yet the neural representations of probability remain elusive. While previous studies have shown that neural activity in several brain regions correlates with probability-related factors such as surprise and uncertainty, similar correlations have not been found for probability. Here, using 7 Tesla functional magnetic resonance imaging, we uncover a representation of the probability of the next event in a sequence within the human dorsolateral prefrontal and intraparietal cortices. Crucially, univariate and multivariate analyses revealed that this representation employs a highly nonlinear code. Tuning curves for probability exhibit selectivity to various probability ranges, while the code for confidence accompanying these estimates is predominantly linear. The diversity of tuning curves we found recommends that future studies move from assuming linear correlates or simple canonical forms of tuning curves to considering richer representations whose benefits remain to be discovered.

Introduction

Our world is conveniently described in terms of probabilities. Examples include perception (e.g. “I’m sure I saw someone I know in the crowd” (Kersten et al., 2004)), social interaction (“My boss is probably lying about my pay rise” (Diaconescu et al., 2014)), and prediction (“It is unlikely to rain tomorrow” (Yang & Shadlen, 2007)). These probabilities reflect the uncertainty of our beliefs about the world, due to the ambiguity of our inputs or limitations in our information processing capabilities (Walker et al., 2023). Human behavior is adapted to these probabilities, as shown by numerous examples both in laboratory experiments and in real life. To name a few, perception is tuned to the probability of occurrence of a given object or feature in the visual, auditory, and somatosensory spaces (Summerfield & de Lange, 2014), and choices are guided by the probability of reward (Rangel et al., 2008). These behavioral results imply that probabilities must be encoded in the brain.

However, the neural code for probabilities remains largely unknown. Several studies have reported that neural activity increases (or decreases) monotonically with quantities that are related to probabilities, but not with probabilities themselves. Examples of related quantities include the coding of prediction error (Schultz et al., 1997), surprise (O’Reilly et al., 2013), and predictability (Bach et al., 2011; Fiorillo, 2003; Monosov & Hikosaka, 2013) . Some studies explicitly reported a failure to identify correlations between probabilities and neural activity (Lebreton et al., 2015; Marshall et al., 2022).

Here, we tested the possibility that neural activity does not scale with probability levels, but instead that the neural code for probability is nonlinear. The codes we studied belong to the broader class of codes that are characterized by tuning curves. A tuning curve relates the value of a quantitative feature (e.g. the orientation of a grating (Hubel & Wiesel, 1959), the number of items in a set (Eger, 2016; Moser et al., 2008; Nieder, 2016)) to the neural activity that this value elicits on average across trials. A special case of tuning curve is the linear code, where neural activity is simply proportional to the encoded quantity. The linear code has been extensively tested and it accounts for the encoding of various quantities such as prediction error, surprise, predictability as mentioned above, and other quantities not related to probabilities, e.g. value (Padoa-Schioppa & Assad, 2008), complexity (Wang et al., 2019), salience (Kutlu et al., 2021). There are also numerous examples of quantities that are encoded with a nonlinear code, such as orientation (Hubel & Wiesel, 1959), numerosity (Nieder, 2016), proportion (Jacob & Nieder, 2009). For these quantities, tuning curves are not linear, but highly non-monotonic. They are often bell-shaped, but more complex tuning curves have also been observed (Diehl et al., 2017; Hardcastle et al., 2017).

To adjudicate between linear and nonlinear codes for probability, it is necessary to characterize the brain’s tuning curves for probability. We measured neural activity with ultra-high field fMRI (7T) because it provides a high signal-to-noise ratio to capture the organization of the neural code at a millimeter-scale with a whole-brain coverage (which is very useful given the lack of anatomical priors on the brain regions involved in the representation of probabilities). We characterized tuning curves for probability in each fMRI voxel using a method that combines the generality of function approximation with basis functions (Bishop, 2007) and linearizing encoding models (Huth et al., 2016; Naselaris et al., 2011). This method does not make assumptions about tuning shape, in contrast to related encoding methods (see Discussion, (Dumoulin & Wandell, 2008; Friston et al., 2007)). It is also more informative about tuning curves than decoding methods. For example, linear classifiers (e.g. regularized linear regression (Findling et al., 2023), support vector machine (Eger et al., 2009)) are often used to decode a quantity from brain signals, but these methods can achieve high performance regardless of whether the underlying tuning curves are linear, bell-shaped, or even more complex (Kriegeskorte & Diedrichsen, 2019). We analyzed the tuning curves for probability both at the single-voxel level and at the level of populations of voxels to strengthen our conclusions. At the single-voxel level, following a univariate approach, we quantified the form and degree of nonlinearity of tuning curves. At the population level, following a multivariate approach, we quantified whether patterns of voxel responses (more specifically their similarity across probabilities) conformed to a linear or a highly nonlinear code.

In terms of experimental design, studying the neural representation of probability requires minimizing potential confounds between probability, and other constructs and task features (Walker et al., 2023). For example, in several previous studies the concept of probabilities has been confounded with evidence for a decision (Yang &

Shadlen, 2007), or with the reward that the subject expects to receive (Ferrari-Toniolo & Schultz, 2023; Kepecs et al., 2008). Probabilities are also correlated with some motor or sensorimotor transformations in many tasks (Yang & Shadlen, 2007), which is useful for studying the behavioral relevance of probabilities, but is a nuisance when studying the neural code of probabilities themselves. To alleviate the problem of confounding factors and variables, we used a probability learning paradigm. Participants estimated the (changing) generative probability of occurrence of neutral items presented sequentially. To limit confounds related to motor actions and decision making, this estimation was covert on most trials (on which the analysis focused), and behavioral reports of probabilities were kept to a minimum. The design had no overt rewards to avoid confounds related to valuation processes. Finally, probability is a latent parameter of this task (i.e., it is not observable in stimulus space), which minimizes confounds related to sensory processes.

To anticipate our result, we provide evidence for the nonlinearity of the neural code of probability, and we strengthen this conclusion by comparing it to another quantity whose code is linear. There is evidence from previous studies that the confidence that accompanies a probability estimate in a learning task correlates (linearly) with neural activity, particularly in regions of the fronto-parietal network (Bounmy et al., 2023; McGuire et al., 2014; Meyniel & Dehaene, 2017; Tomov et al., 2020). The comparison between probability and confidence is not confounded by between-subject differences, or differences in perceptual space since both are derived from the very same sequence of stimuli.

Results

Task probing the representation of probabilities

We subjected twenty-six human subjects to a probability learning task and simultaneously measured their brain activity using ultra-high field (7T) fMRI, to examine their neural representation of probabilities, and of the confidence associated with their probability estimates. After one training session out of the scanner, subjects performed four sessions of the task in the MRI scanner. In each session, subjects observe a sequence of 420 stimuli appearing one by one, each with a binary value (A or B) sampled from a hidden generative probability p(A) (Fig. 1). This hidden probability undergoes abrupt change points at random unpredictable times which were not signaled to the subjects. The subject's goal is to estimate the current value of the hidden probability throughout the sequence. To perform the task correctly, subjects have to frequently update their estimate as new observations are made. This experimental protocol allows us to efficiently probe the neural representation of probabilities and confidence as it induces frequent variations in these quantities that can be compared with variations in measured neural activity.

To minimize disruption to the fMRI signals, we asked subjects to perform their estimation covertly and only occasionally requested behavioral reports. The reports occurred during dedicated periods (on average every 22 stimuli), during which the subject had to choose a range for their current probability estimate and associate a level of confidence with that estimate. We found a strong correspondence between the estimates reported by the subjects and those of the normative model (see Methods) (Fig. 1 B; Pearson r=0.79 mean ±0.12 s.e.m, t_{25}=31.8, $p<10^{-21}$ for probability and r=0.16±0.13, t_{25}=5.9, $p<10^{-5}$ for confidence). In another recent study, we found that this strong correspondence with the normative model was consistently observed at the trial-by-trial level (Foucault & Meyniel, 2023). In what follows, we thus used the estimates of the normative model as a proxy for those of the subjects in order to relate trial-by-trial estimates with ongoing neural activity.

Encoding models and evaluation procedure

To relate probability and confidence estimates to neural activity, we constructed models that predict neural activity as a function of the estimate of interest (probability or confidence), which we refer to as encoding models (Fig. 2A). We defined two classes of models: the class of linear encoding models, which is the one classically used in fMRI studies and which assumes a linear relationship between the variable of interest and neural activity, and the class of versatile encoding models, which does not make such an assumption. Instead, the versatile class captures all kinds of relationships, which include but are not restricted to linear relationships, by leveraging approximation theory. The versatile class can approximate arbitrary functions as a weighted sum of basis functions (see Methods).

To evaluate the encoding models, we incorporated them into a larger pipeline to obtain predictions at the level of the fMRI signal measured in each voxel (Fig. 2A). The resulting model predicts the fMRI time series in a given session from the probability or confidence estimates, as well as several factors of no-interest that we aim to control for, including: the stimulus onset, the surprise elicited by the stimulus, the entropy associated with the probability estimate (a measure of unpredictability), factors related to the report periods, and motion factors.

We fitted the model parameters and tested the fitted model on independent data using a leave-one-session-out cross-validation procedure (see Methods). During testing, we only kept the part of the model corresponding to the factor of interest (probability or confidence) and calculated the R^2 score, which represents the predictive accuracy of the model. Finally, to ensure that a positive score can only be obtained by an encoding related to the sequence specifically observed by the subject and not merely to the statistical structure of the sequences in general, we calculated a null distribution of R^2 scores by performing the same analysis after replacing the true sequence with other sequences, and standardized the score obtained with the true sequence relative to this null distribution to yield the final score, which we call z-R^2.

We validated our approach end-to-end using simulations (Fig. 2B). By simulating an experiment following our protocol under the hypothesis that neural activity noisily encodes one of the two types of estimate (probability or confidence) according to an encoding belonging to one of the two classes (linear or versatile), and by applying our fMRI analysis pipeline to the simulated fMRI signal, we established that when neural activity encodes the same type of estimate as the model, the linear model only explains the signal well when this encoding is linear, while the versatile model explains the signal well no matter whether this encoding is linear or not. When neural activity and the model encode a different type of estimate, the model does not explain the signal.

Neural encoding of probability

The versatile model revealed a significant encoding of probabilities in the prefrontal and parietal cortex (Fig. 3 and Table 1). In contrast, the linear model did not reveal any encoding of probabilities across the cortex, even below the significance threshold (see Fig. 3). The superiority of the versatile encoding model over the linear one is confirmed by the statistical map of the difference in R^2 the two models (Supplementary Fig. 1).

This pattern of results is consistent with those predicted under the hypothesis that neural activity encodes probabilities in a nonlinear way (Fig. 2B, second row and first two columns of the matrix). Note that these results do not depend much on the specific choice of basis functions, provided that they have the same approximation properties (see Supplementary Fig. 2).

Neural encoding of confidence

Contrary to probabilities, for confidence, the linear encoding model significantly explains the measured signal in large regions around the intraparietal and precentral sulci (Fig. 4 and Supplementary Table 1). This is consistent with previous studies on the neural correlates of confidence (Bounmy et al., 2023; Meyniel & Dehaene, 2017). As expected, the versatile encoding model also captures the signal in the regions explained by the linear model. However, unlike probabilities, the regions captured by the versatile and the linear encoding models are very similar throughout the cortex (Fig. 4). This indicates that probability and confidence are encoded differently by the brain, which we sought to further characterize next.

Univariate characterization of the neural code

To characterize the neural code, we reconstructed the tuning curves measured at the vertex level. These curves were obtained by calculating the sum of the basis functions weighted by the fitted weights of the versatile model (Fig. 5A). For each subject, we focused on a set of vertices where the predictive accuracy of the versatile model was large enough to ensure a reliable characterization of the tuning curves, which we verified by estimating the tuning curves independently on two halves of the data (Pearson correlation of the estimated weights on the two halves for the vertices of interest: 0.55±0.03 and 0.51±0.04 for probability and confidence, respectively). See examples of tuning curves estimated on the whole and on two halves of the data in Fig. 5B.

We used three characteristic measures to describe and compare the tuning curves for probability and confidence (Fig. 5C). For all three measures, the neural encoding of probability and confidence differed significantly. The first measure looked at the location of the maximum of the tuning curve (i.e., the probability or confidence value that maximizes neural activity). This maximum is expected to be close to one of the two extremes in the case of a linear code. Compared to confidence, tuning curves for probability were more often maximized at non-extreme values (Fig. 5C, proportion: 84±6 % for probability vs. 42±6 % for confidence, $p<0.001$, two-tailed t-test). The second measure looked at the degree of non-monotonicity of the tuning curves, quantified by a continuous index between 0 and 1 (see illustration in Fig. 5C). According to this index, tuning curves were more non-monotonic for probability than for confidence (Fig. 5C, 0.61±0.04 for probability vs. 0.36±0.04 for confidence, $p<0.001$, two-tailed t-test). The third measure looked at the degree of nonlinearity of the tuning curve (corresponding to the degree of a polynomial, the higher the degree, the more nonlinearities). This measure showed that tuning curves were more nonlinear for probability than for confidence (Fig. 5C, 4.5±0.4 for probability vs. 3.5±0.2 for confidence, $p<0.05$, two-tailed Kruskal-Wallis test by ranks).

Multivariate characterization of the neural code

So far, the analysis has been univariate, focusing on one voxel at a time. If different voxels are tuned to different ranges of probability, as indicated by the above encoding analysis, then patterns of voxel responses should be informative about probability, which we tested with a multivariate decoding approach. Furthermore, the pattern of voxel response should exhibit different geometric properties if they arise from nonlinear vs. linear codes, providing another test of the hypothesis that the neural code for probability is highly nonlinear. Below, we present thise two analyses.

We measured the extent to which probability (and for comparison, confidence) could be decoded in different brain regions by dividing the cortex following the parcellation by Glasser et al (Glasser et al., 2016) and pooling homologue regions in both hemispheres, resulting in 180 regions. We based our decoding approach on the versatile encoding model presented above. The versatile encoding model quantifies in each voxel the weights of basis functions of probability. The basis functions being equally spaced narrow Gaussian functions, the set of estimated weights can be thought of as characterizing the voxel responses to different bins (i.e. narrow ranges) of probability. We tested whether it is possible to identify (i.e. decode) these probability bins given the pattern of voxel responses that they elicit. We used 5 probability bins instead of 10 as in the encoding approach presented above to make the estimates of weights more reliable, and thus decoding easier. We adopted the same approach to decode confidence levels.

We trained and tested the decoder on different data sets, using leave-one-session-out cross validation at the subject-level (Varoquaux et al., 2017). First, we reduced dimensionality in each region by selecting the 100 voxels that are the most informative about probabilities (using the $z\text{-}R^2$ metric introduced above, but estimated using the training sessions only). Then we estimated the patterns of voxel responses for different probability bins, for the test session on the one hand and the three training sessions together on the other hand. Last, we decoded the probability bin corresponding to a response pattern in the test session by identifying the probability bin eliciting the most similar response pattern in the training session. We computed the decoding accuracy for each brain region and tested for statistical significance against chance-level accuracy at the group level (see Fig. 5 and Methods).

Decoding accuracy was generally larger for confidence than probability (Fig 6, Tables 2 & 3), which is expected given that the versatile encoding model accounted for voxel responses in more regions based on confidence than on probability (Fig. 4 vs. 5). For probability, we found 6 regions with FDR-significant decoding accuracy. They comprised the dorsolateral prefrontal cortex and the intraparietal cortex, which had been identified with the encoding approach. For confidence, more regions exhibited a FDR-significant decoding accuracy, notably in parietal and prefrontal cortex, similar to the regions identified with the encoding approach.

Decoding accuracy indicates that patterns of voxel activity are informative about probability or confidence, but it does not characterize the type of neural code being used. We examined more closely the patterns of voxel responses to test for the existence of different codes for probability and confidence. More precisely, we examined the matrices that quantify the dissimilarity of patterns of voxel responses between bins of probability (and similarly,

bins of confidence) that served as a basis for decoding. These matrices are called representational dissimilarity matrices (RDM) (Kriegeskorte & Kievit, 2013). Different types of code predict different types of RDM. If the code is highly nonlinear and non-monotonic, as suggested by the encoding analysis of probability, then the patterns of voxel responses should be maximally similar when representing the same probability bin, and equally dissimilar between a given bin and any other bin, resulting in an identity RDM (Fig. 7A). In contrast, if the code is highly linear (or more generally monotonic), as suggested by the encoding analysis of confidence, then the patterns of voxel responses should be more similar for confidence bins that are closer to one another, resulting in a graded RDM. To determine which code best accounted for the empirical RDMs, we regressed the RDMs for probability (and confidence) obtained in each region onto the identity RDM and graded RDM arising from highly nonlinear and linear codes, respectively.

We first focused on the 5 regions exhibiting the most significant decoding accuracy, separately for probability (Fig. 7B) and confidence (Fig. 7C). The identity RDM significantly accounted for the RDMs for probability, while no significant effect of the graded RDM was found, and the difference between the two approached significance in most regions. In contrast, for confidence, the graded RDM significantly accounted for the RDMs, the identity RDM yielded no significant effect, and their difference was significant in most regions.

We then carried another analysis that covered all regions. We counted the number of regions best explained by the identity or the graded RDM based on the maximum regression coefficient. A majority of regions followed the graded RDM in the case of confidence (M_{graded}=0.672, SEM=0.132, t=3.6, p=0.0014, t-test against 0.5) whereas a majority of regions followed the identity RDM in the case of probability (M_{graded}=0.436, SEM=0.086, t=-1.8, p=0.078, t-test against 0.5), and the difference between confidence and probability was significant (paired difference 0.236, SEM=0.0462, t=3.6, p=0.0014, t-test). Together, these results indicate that the codes for probability and confidence are different, being respectively highly nonlinear and mostly linear, which confirms the conclusion of the encoding analysis presented above.

Discussion

We used a task in which participants estimated the latent probability of a stimulus occurring in a sequence. Subjects accurately tracked this latent probability, as revealed by the comparison of their reports with an ideal observer. We identified a representation of this latent probability in the fMRI signals recorded outside of the report periods, particularly in the dorsolateral prefrontal cortex and intraparietal cortex. Crucially, this representation was based on a nonlinear code. In contrast, the representation of the confidence that accompanied the probability estimate was based on a linear code. Detailed analysis revealed that the vast majority of fMRI voxel tuning curves for probability were non-monotonic, with one or more local extrema. In contrast, the tuning curves for confidence were essentially linear. In addition, a linear code and a highly nonlinear code are expected to result in different geometries, which we confirmed with a multivariate analysis of the patterns of voxels responses to probability and confidence.

Our results relied on the use of a versatile encoding model capable of accommodating tuning curves of almost any shape. To this end, we combined the universal function approximation properties of basis sets (Bishop, 2007; Franke, 1982) with the use of linearizing encoding models for fMRI (Huth et al., 2016; Naselaris et al., 2011). This method consists of applying basis functions to a quantity of interest to obtain features and then modelling the fMRI signal in each voxel as a linear combination of these features, following the general linear model approach that is massively used in fMRI (Friston et al., 2007). This approach is related to other methods that can accommodate nonlinear tuning curves, such as population receptive field (pRF) mapping (Barretto-García et al., 2023; Dumoulin & Wandell, 2008; Harvey et al., 2013). The key difference is that pRF methods assume a specific form of nonlinearity, typically bell-shaped tuning curves, corresponding to the idea that a voxel is selective for a range of values. In contrast, our method can accommodate tuning curves exhibiting a selectivity for multiple value ranges. It turns out that approximately a third of tuning curves we obtained for probability did not conform to a bell-shaped tuning as they exhibited more than one peak. Our finding that the neural representation of probability is highly nonlinear retrospectively explains null findings in previous studies that used methods assuming a linear code (Bounmy et al., 2023; Lebreton et al., 2015; Marshall et al., 2022).

Our results raise the puzzling question of why some quantities are encoded with linear codes, such as confidence here, or reward (Lebreton et al., 2009; Padoa-Schioppa & Assad, 2008), salience (Kutlu et al., 2021), surprise (Meyniel & Dehaene, 2017; O'Reilly et al., 2013), prediction error (Pessiglione et al., 2006; Schultz et al., 1997), evidence accumulation (Brunton et al., 2013; Gold & Shadlen, 2007) and some other quantities are encoded with nonlinear codes, such as probability here, or orientation of visual object (Hubel & Wiesel, 1959), angle in arm reaching (Georgopoulos et al., 1986), numerosity (Nieder, 2016). This question is beyond the scope of our study, but we mention some speculations. Some scalar quantities lie on an axis whose direction is relevant to the regulation of behavior and brain processes. For instance, humans and other animals generally seek more, not less, rewards (Rangel et al., 2008). A linear code with increasing activity levels for larger rewards may, in downstream circuits, facilitate the invigoration of behavior to obtain larger rewards (Pessiglione et al., 2007), and the comparison of different reward levels (which, in a linear code, simply amounts to comparing activity levels). A similar argument applies to salience, surprise, accumulated evidence, and lack of confidence. Higher values of these quantities usually enhance other processing: more salient and surprising events elicit stronger orienting responses (Sara & Bouret, 2012), lower confidence about a learned estimate increases the learning rate (Foucault & Meyniel, 2023; Nassar et al., 2010). In contrast, in our task, the probability of occurrence of a right- vs. left-tilted Gabor patch has no valence and it is not immediately relevant to behavior. Interestingly, the average response across a population (of neurons or voxels) is invariant to the value encoded in a nonlinear code with sufficiently diverse tuning curves. Assuming that more neural activity is costly (Gallistel, 2017), this invariance property implies that the same energy budget is expended to represent any probability, in particular for low (close to 0) and large (close to 1) probabilities. In contrast, encoding reward, salience, surprise, or lack of confidence with increasing (linear) levels of activity appropriately expends more energy on larger, behaviorally relevant values.

Here, we found a neural representation of probabilities predominantly in the dorsolateral prefrontal cortex, the precentral sulcus and the parietal region. The dorsolateral prefrontal and intraparietal cortices have been reported to host a general coding system for magnitudes of different types, from number of objects to proportions in humans and monkeys (Eger, 2016; Nieder, 2016). Our results suggest that this general coding system may also encode probability. In our results, the dorsolateral portion of the prefrontal cortex appeared to encode only probability, whereas the precentral sulcus and the parietal regions encoded both confidence (with a linear code) and probability (with a nonlinear code). Simulations showed that a linear code for confidence and a nonlinear code for probability can be clearly distinguished from one another; their co-localization is therefore a notable finding. We speculate that if a region is involved in estimating and encoding of probabilities, it should be more active when the estimate is updated more, which typically occurs when confidence is lower (Bounmy et al., 2023; McGuire et al., 2014; Tomov et al., 2020). Co-localization of a nonlinear code for probability and a decreasing (linear) code for confidence would then be expected. Future studies should distinguish between updating and representing probabilities, and our results suggest that these processes may differentially involve the dorsolateral prefrontal cortex on the one hand, and the precentral sulcus and parietal region on the other hand.

It is important to distinguish between the representation of an event probability and the representation of a probability density function (of scalar variables such as the orientation of a grating). The latter has been investigated by two lines of research focusing on encoding and decoding, respectively. On the one hand, some researchers have proposed that probability density functions are encoded in activity patterns (Jazayeri & Movshon, 2006; Zemel et al., 1998). On the other hand, research on probabilistic population codes (Ma et al., 2006) posits that, with certain models of neural variability, probability density functions can be decoded from activity patterns in the form of a posterior distribution over some scalar variable. Here, we have identified a neural code for the event probability (which is the mean of the posterior distribution); it remains open whether the brain also codes a probability density function of the event probability (i.e. its full posterior distribution, see Eq. 1 in Methods). Following the encoding approach (à la Zemel et al), we have tested a variant of the versatile encoding model that encodes the full posterior distribution of the event probability (instead of its mean; see Methods). We found no clear evidence for such a representation (see Supplementary Fig. 3). We did not attempt to follow the probabilistic population code approach (à la Ma et al), because it requires modelling the variability of the responses elicited by a given probability (van Bergen et al., 2015), which is difficult to obtain in practice but promising for future research.

We acknowledge that our results do not exclude the possibility that other coding schemes are used to represent probability. Functional MRI can study rate codes (in which information is conveyed by the rate of spikes, not their timing) at the level of voxels, but its temporal resolution is incompatible with the study of temporal codes (Dayan & Abbott, 2005), especially at the level of neurons. Temporal codes may also be used for probabilities. Probabilities could in principle also be stored in synaptic weights (Iigaya, 2016) or intracellular substrates (Gallistel, 2017). It has been claimed that the function of neural activity (and thus indirectly, the fMRI signal (Logothetis et al., 2001)) is to transmit information, which is much more energetically costly than storing information in a cellular substrate (Gallistel, 2017). We note that the probability here has to be constantly updated because there are change points occurring at random, un-signalled times. The reason why we found a representation of probabilities in fMRI activity patterns may be because probabilities were being constantly updated by subjects. This updating process, which involves the interaction of neurons that combine the current estimate with information about the incoming stimulus, may cause fMRI activity patterns tuned to probability. In this view, had the probabilities not been frequently updated, we may not have detected them in activity patterns. This possibility remains to be tested.

We now turn to discuss some limitations of our approach. First, we have used binary sequences with stimulus A or B, so that the generative process can be described equivalently in terms of p(A) or p(B), since p(A)=1-p(B). We found evidence that some pieces of the cortex exhibit similar activity for values of p(A) that are symmetric with respect to 0.5, e.g. when p(A)=0.2 and p(A)=0.8 (see Supplementary Fig. 4). This may be due to a representation of the event probability switching between p(A) and p(B), or due to a similar proportion within a voxel of neurons each coding for either p(A) or p(B). Another possibility is that some representations of probability may focus on the probability of the most likely stimulus (the one with p>0.5) and the identity of this stimulus (these two pieces of information are sufficient to reconstruct p(A) and p(B) in the binary case). The use of more than 2 items in a sequence would be useful to adjudicate between these different possibilities.

Invariance is a useful criterion for testing for neural representations, especially in the context of uncertainty (Walker et al., 2023). The invariance of probability representation could be tested with respect to the features of the stimulus (e.g. visual stimuli that differ in shape, or color, rather than orientation) or the sensory modality used (e.g. auditory stimulus (Meyniel & Dehaene, 2017)). The origin of the probability could also be changed. Here, the probability originates from a statistical learning process operating on a sequence, but not all probabilities do. Reasoning and memory can also be used, for example when estimating the probability that a statement is true, such as “Is Paris bigger than Berlin?” (Lebreton et al., 2015).

In addition, the timing of the task and the poor temporal resolution of fMRI precluded the use of trial-level decoding, which was done at the session level here. Future studies could explore a decoding of probability across trials, perhaps benefiting from the use of better time-resolved recordings such as electrophysiology.

In summary, we have unraveled a neural representation of event probability in the human cortex that is based on a highly nonlinear code, and that cannot be detected by simpler methods that assume a linear code. The methods we have developed here can be used to search for neural representations of probability in many different types of tasks.

Methods

Participants and task

The behavioral and fMRI data used in this paper have already been analyzed in (Bounmy et al., 2023). Experimental protocols were approved by the local ethics committee (CPP-100032 and CPP-100055 Ile-de-France) and the informed consent of all 29 participants (15 female, mean age 25.4 ±1.0 s.e.m.) was obtained before they began the experiment. Three participants were excluded from analysis due to acquisition problems, resulting in an effective total of N=26 subjects for analysis.

After receiving task instructions and completing a training session, participants entered the MRI scanner and performed 4 task sessions, during which the scanner recorded functional MRI data. Each session lasted

approximately 11 minutes and consisted of a sequence of 420 stimuli presented one after the other, for 1 s each, with an inter-stimulus interval of 0.3 s (see example in Fig. 1A).

Each stimulus had a binary value, A or B, represented in the task by one of two distinct orientations of a high-contrast grating, easily distinguishable from each other. The values were drawn from a Bernoulli distribution with a hidden generative probability p(A) to be learned, whose value will be denoted h_t. The hidden probability h_t followed a stochastic change-point process: at each time step, it could either remain the same or undergo a change point, with a change-point probability of 1/75 under the constraint of a maximum period of 300 stimuli without change points. The value of h_t was uniformly sampled between 0.1 and 0.9 initially and at each change point, under the constraint that the odds of A changed by a factor of at least four.

Throughout the task, the subjects' goal was to estimate the hidden probability h_t. The occurrence of unsignaled and unpredictable change points made the estimation all the more challenging. Accurate estimation requires making probabilistic inferences about the value of h_t given the observed stimuli, with knowledge of the generative process of the sequences, as described in the task model below. Subjects could make such inferences as they had been briefed about the generative process in an informal way during the instructions phase.

Subjects were instructed to continuously estimate the probability during the sequence, and were occasionally asked to report their estimate during dedicated time periods. Isolating the report periods from the periods of stimuli and the estimate updates they elicited allowed us to ensure that neural representation of the factors of interest we found in the fMRI recordings, related to the estimation process, were not confounded by the subject's reporting. Report periods occurred on average every 22 stimuli, with a random uniform jitter of ±3 stimuli maximum. Each report period consisted of a response screen where the subject made two choices: a choice of estimate range for h_t among three or five possible ranges (the scale was randomly selected between the three-choice scale and the five-choice scale for each period, except for the first half of participants where the scale was always the five-choice scale), and a choice of confidence level associated with this estimate among five possible levels. These choices were made using a right-handed five-finger button pad. Although the choices were discrete, subjects were instructed to internally generate continuous-valued estimates. They were further incited to do so in order to produce correct reports as they did not know in advance which scale, and therefore which choice ranges would be available (there being no trivial mapping between the two scales).

Task model

The model we used for the task is the normative model, in the sense that it produces, for each trial, the optimal estimate of the hidden probability that the subject could have produced, given the stimuli they have observed so far and their knowledge of the generative process of the task. This estimation problem requires inferring the value of h_t given the stimulus values observed in the past $s_{1:t}$. Since the generative process is probabilistic, the inference problem is also probabilistic (the value of h_t cannot be determined with certainty). Using the rules of probability calculus and in particular Bayes' rule, a posterior distribution on h_t can be calculated, denoted as $p(h_t \mid s_{1:t})$.

In the present case, one solution to calculate algorithmically the posterior distribution is to proceed iteratively on the observations, initializing $p(h_0)$ to a uniform distribution and computing $p(h_{t+1} \mid s_{1:t+1})$ based on the value of s_{t+1} and the previously computed $p(h_t \mid s_{1:t})$. This calculation is done using the following formula.

$$p(h_{t+1} \mid s_{1:t+1}) \propto p(s_{t+1} \mid h_{t+1}) \int p(h_{t+1} \mid h_t)\, p(h_t \mid s_{1:t})\, dh_t \quad [1]$$

This formula is derived from the rules of probability calculus by leveraging two conditional independence properties of the present generative process: (1) s_{t+1} is conditionally independent of $s_{1:t}$ given h_{t+1}; (2) h_{t+1} is conditionally independent of $s_{1:t}$ given h_t. See Supplementary Note 1 for the derivation.

The $\propto$ operator denotes the equality up to a constant factor. This constant is implicitly factored in at computation time by normalizing the right-hand side of equation [1] so that it sums up to 1 over the possible values of h_{t+1}, to obtain the left-hand side. The term $p(s_{t+1} \mid h_{t+1})$ is given by the Bernoulli distribution, equal to h_{t+1} or $(1-h_{t+1})$

depending on whether s_{t+1} is A or B. The term $p(h_{t+1} \mid h_t)$ reflects the generative probability that a change point occurs (1/75) or does not occur (74/75), depending on whether h_{t+1} is different from or equal to h_t, respectively.

The normative estimate of hidden probability and the associated confidence are both calculated from the posterior distribution. The probability estimate is equal to the mean of the posterior, $E[h_{t+1} \mid s_{1:t}]$. It is optimal in the sense that it minimizes the mean squared error with the true value, and is equal to the posterior probability that the next stimulus value is A (this is formally what subjects were asked to report). Confidence was defined as $-log\ SD[h_{t+1} \mid s_{1:t}]$, the log-precision of the posterior (up to a factor of two). Hereafter, we will use the symbol x to refer to these two types of estimates indiscriminately (x_t representing the estimate for stimulus s_t) as the presented encoding models are mostly independent of the specific type of estimate being encoded.

The task model was implemented using Python and the NumPy package (https://numpy.org).

Encoding models

Encoding models predict the fMRI activity of a voxel as a function of the estimates obtained from the stimulus sequence seen by the subject. We defined four main encoding models, 2×2, depending on whether the encoded estimates are probability or confidence estimates, and whether the model tuning curve function belongs to the linear or the versatile class (Fig. 2).

We also considered another model, following a hypothesis proposed in the literature, in which the activity was a function of the entire posterior distribution (rather than a moment of the distribution, like the probability and confidence estimates are) (Sahani & Dayan, 2003; Zemel et al., 1998). That is, in that model, $f_i(x)$ in equation [3] was replaced by its posterior mean $\int f_i(x)p(x)dx$, where p is the posterior distribution. However, when we tested that model on our fMRI data, it explained the data less well than the simpler models encoding the probability or confidence estimates (see Supplementary Fig. 3). Therefore, we focused on the simpler models.

The probability and confidence estimates are calculated from the sequence as explained in the "Task Model" section above. The model tuning curve function maps the encoded estimate, x, to a prediction of neural activity, $\hat{y}$, and is parameterized with weights w that are to be fitted to the data. In the linear class, this function is of the form:

$$\hat{y} = w\ x \quad [2]$$

In the versatile class, this function is of the form:

$$\hat{y} = \sum_{i=1}^{K} w_i\ f_i(x) \quad [3]$$

where the f_i are (radial) basis functions that have approximation properties (Franke, 1982). Here, we used Gaussian basis functions, but we also performed simulations using sigmoid basis functions, and the results were similar whether we used Gaussian or sigmoid functions (Supplementary Fig. 2). The Gaussian basis functions are expressed $f_i(x) = c\ exp[-(x-\mu_i)^2 / (2\sigma^2)]$. The centers of the basis functions μ_i were distributed to have equal spacing between two consecutive centers, between the lower bound of the interval and the first center, and between the upper bound and the last center (the interval being [0, 1] for probability, and [1.1, 2.6] for confidence). The number of basis functions was K=10, and their dispersion was σ=0.04 for probability and σ=0.06 for confidence. These values were determined through simulations to optimize the R^2 scores averaged over a wide range of simulated activity with different K and σ values.

To transform the predictions of the theoretical models described above into predictions at the level of fMRI activity in a voxel, we convolved the encoding model regressors (that is, the scalar quantities that are multiplied with the weights w: x in the linear case and the $f_i(x)$ in the versatile case), with the canonical hemodynamic response function at the onsets of the corresponding stimuli (these convolved regressors are often referred to as "parametric modulations" in the fMRI community).

Additionally, we included in the encoding model other regressors corresponding to factors of no-interest in this study, which we removed during testing to evaluate the specific encoding of probability or confidence. The regressors of no-interest were the following.

- Parametric modulations of stimulus onsets associated with factors other than probability and confidence:
  - a constant
  - the Shannon surprise induced by the stimulus, *-log p(s)*, where *s* is the stimulus value and *p(s)* is the normative probability estimate for that stimulus value given the previously observed stimuli
  - the Shannon entropy of the outcome implied the normative probability estimate, *-p log(p) - (1-p) log(1-p)* (see Supplementary Fig. 4).
- Parametric modulations of response screen onsets modeling reporting periods (including, for each period, a response screen for probability and another for confidence):
  - a constant
  - the normative estimate
  - the estimate reported by the subject
- Six motion regressors

Finally, we applied the same temporal preprocessing to all regressors as we applied to the fMRI signal (detrending, filtering, z-scoring across sessions, and session-wise demeaning, see MRI data preprocessing section).

At the voxel level, the fMRI activity predicted by the model after preprocessing, $\hat{y}_{fMRI}$, is written as $\hat{y}_{fMRI} = w\,\dot{x} + \boldsymbol{w_n}\,\boldsymbol{n}$ for the linear model, and $\hat{y}_{fMRI} = \boldsymbol{w}\,\dot{\boldsymbol{f}}(x) + \boldsymbol{w_n}\,\boldsymbol{n}$ for the versatile model, where $\dot{}$ represents the convolution and temporal preprocessing operations, $\boldsymbol{w}$ and $\boldsymbol{f}$ are the vectors $[w_1, \ldots, w_k]$ and $[f_1, \ldots, f_k]$, $\boldsymbol{n}$ is the vector comprising all preprocessed regressors of no-interest, and $\mathbf{w_n}$ is their associated weights vector (bold symbols denote vectors as opposed to scalar quantities).

The encoding models were implemented using Python, NumPy (https://numpy.org), and nilearn (https://nilearn.github.io).

Evaluation of the encoding models

We evaluated the ability of the encoding models to predict fMRI data for a given subject using leave-one-session-out cross-validation: three sessions were used for training the model, and the fourth, left-out session, was used to test the trained model. This procedure was repeated for each of the four possible choices of the left-out session, and the test scores obtained were averaged across the four left-out sessions.

Training. During training, the weights of the encoding model were fitted using Ridge regression. The Ridge penalty was calculated using an analytical formula that adjusted the amount of penalty (λ) based on the number of model parameters (m): $\lambda = 199m$, which was validated through simulations.

Testing. During testing, we calculated the fMRI activity predicted by the model using only the part of the model associated with the factors of interest (probability or confidence). (This is equivalent to replacing the regressors of no-interest with their mean value for the session, which is equal to 0 after session-wise demeaning.) As a score, we first calculated the R^2 obtained by comparing the model predictions with the actual fMRI data, using the sums-of-squares formulation (Poldrack et al., 2020). We then calculated a null distribution of R^2 scores by injecting null predictions into the R^2 calculation. These null predictions were obtained by replacing the true stimulus sequence seen by the subject with another sequence randomly generated according to the task generative process, for 100 generated sequences. Finally, we calculated the score we call $z\text{-}R^2$ by standardizing the R^2 score obtained for the true sequence by the mean, $\mu_0(R^2)$, and standard deviation, $\sigma_0(R^2)$, of the null distribution: $z\text{-}R^2 = (R^2 - \mu_0(R^2)) / \sigma_0(R^2)$.

Models were fitted and tested using Scikit-Learn (https://scikit-learn.org).

Simulation of encoding models

We conducted simulations to verify that our procedure, applied in our experimental protocol, was able to detect and differentiate a neural encoding of probabilities or confidence, linear or nonlinear. For this purpose, experimental data was generated assuming a certain encoding model, and the generated experimental data was analyzed with other encoding models. As for subjects, each generated experiment consisted of four sessions, with one sequence

of stimuli per session, generated according to the task process. Noisy fMRI activities were then generated for each session and a certain number of simulated voxels assuming a certain generative model of neural activity, which corresponded to one of the four encoding models presented above (encoding of probability or confidence estimates, according to a linear or a versatile encoding model).

The procedure for generating fMRI activity was as follows. For each simulated voxel, we randomly generated weights for the generative model by drawing each weight uniformly in [-0.5, 0.5]. We generated the "signal" component of the fMRI activity in accordance with the generative model, by calculating the weighted sum of the corresponding fMRI regressors as described in the above section on encoding models. We then injected Gaussian white noise with power equal to nine times that of the signal, in order to obtain a signal-to-noise ratio of 10% signal to 90% noise. This produced a set of fMRI activities for each generated experiment and each possible generative model.

For each generated experimental data, the procedure presented in the above section was used to fit the encoding models and evaluate their ability to predict the simulated fMRI data (as subsequently done for the subjects' fMRI data). This produced one *z*-R^2 score per voxel, fitted model, generative model, and generated experiment. We averaged the scores obtained across one hundred experiments and one hundred simulated voxels to obtain an average *z*-R^2 score for each possible generative model × fitted model pair, resulting in a 4×4 matrix, shown in Fig. 2B.

This analysis was also performed by splitting the versatile model into one with Gaussian basis functions and one with sigmoid basis functions to produce the matrix shown in Supplementary Fig. 2.

The simulations were implemented using Python and Numpy.

MRI data acquisition

Equipment. The MRI scanner was a Siemens MAGNETOM 7 Tesla at the NeuroSpin center (CEA Saclay, France), with whole-body gradient and 32-channel head coil by Nova Medical.

Functional MRI acquisition. Whole-brain functional volumes with 1.5mm isotropic voxels and T2*-weighted fat-saturation images were acquired using a multi-band accelerated echo-planar imaging sequence (Moeller et al., 2010; https://www.cmrr.umn.edu/multiband/). The sequence parameters were: multi-band factor = 2 (MB); GRAPPA acceleration factor = 2 (IPAT); partial Fourier = 7/8 (PF); matrix = 130 × 130, number of slices = 68, slice thickness = 1.5 mm, repetition time = 2 s (TR); echo time = 22 ms (TE); echo spacing = 0.71 ms (ES); flip angle = 68° (FA); bandwidth = 1832 Hz/px (BW); phase-encoding direction: anterior to posterior.

Before each session, two single-band functional volumes were acquired with the above parameters except that they had opposite phase-encoding directions. This was later used for distortion correction (see below in *Preprocessing*).

A Gradient Recalled Echo (GRE) sequence was acquired for calibration before starting the sessions. For shimming, a B_0 map was acquired and loaded in the console, first of the whole brain, then in an interactive fashion on the occipito-parietal cortex. This aimed to reduce the FWHM and increase the T2*. A B_1 map was then acquired and the intraparietal sulcus values were used to compute a reference voltage from which the system voltage was chosen.

Anatomical MRI acquisition. Whole-brain T1-weighted anatomical images with 0.75 mm isotropic resolution were acquired using an MP2RAGE sequence. The sequence parameters were: GRAPPA acceleration factor = 3 (IPAT), partial Fourier = 6/8 (PF), matrix = 281 × 300, repetition time = 6 s (TR), echo time = 2.96 ms (TE), first inversion time = 800 ms (TI_1), second inversion time = 2700 ms (TI_2), flip angle 1 = 4º (FA_1), flip angle 2 = 5º (FA_2), bandwidth = 240 Hz/px (BW).

MRI data preprocessing

The functional volume slices were corrected for slice-timing with respect to the slice acquired in the temporal middle of a volume acquisition. The functional volumes were corrected for motion using rigid transformations, and co-registered to the anatomical image. Session-wise distortion correction was applied to the functional volumes using FSL apply_topup, after having estimated a set of field coefficients for the session using FSL TOPUP with the two-single band volumes with opposite phase-encoding directions acquired before that session.

Prior to encoding and decoding analyses, the fMRI time series of each voxel and session were temporally detrended and high-pass filtered at a cutoff frequency of 1/128 Hz. The series from the four sessions were then concatenated in order to z-score the fMRI data across the four sessions for each voxel for numerical convenience. Note that we did not z-score the fMRI data per session because under the versatile encoding model, it is expected that the variance of the fMRI signal should change from session to session depending on the probabilities and confidence levels estimated during each session. Finally, the mean of each session was subtracted from the data in order to ignore any changes in the signal baseline between sessions that might be caused by nuisance factors.

Slice-timing correction, motion correction, and co-registration were done using tools from SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12). Distortion correction was done using tools from FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL). SPM12 and FSL were called from Python using the NiPype module (https://nipype.readthedocs.io). Further preprocessing was done using Python and the NumPy, SciPy (https://scipy.org) and nilearn (https://nilearn.github.io) packages.

Conversion of volumetric data into cortical surface data.

Here we detail the processing steps we used throughout this study to project subject-level volumetric data, such as the R^2 and z-R^2 scores computed from the fMRI data (see section on encoding model evaluation), onto the cortical surface, and to bring them into a common space across subjects. These processing steps were performed using FreeSurfer (https://surfer.nmr.mgh.harvard.edu) and the Python interface to FreeSurfer commands provided by the NiPype module. The resulting surface data were then analyzed by working directly with the numerical arrays in Python, unless otherwise stated.

For each subject, the cortical surface was reconstructed from the acquired high-resolution anatomical MRI image by running the FreeSurfer command 'recon-all'. These reconstruction data were then used to project other volumetric data calculated from the functional data onto the cortical surface, working in the subject's native space. This projection was performed using the 'mri_vol2surf' command. The surface data were then normalized, i.e., resampled to be brought into a common space: the 'fsaverage' template of FreeSurfer, version 7 high-resolution (163,842 vertices per hemisphere), and were spatially smoothed with a Gaussian kernel of 3mm full width at half maximum. The resulting data are surface maps containing one data point per vertex (a vertex is the surface equivalent of a voxel in the volume, called vertex because vertices are assembled to define polygons that together form a three-dimensional mesh of the cortical surface).

Group-level statistical maps

The subjects' z-R^2 surface maps were grouped together and a one-sample t-test against zero was performed at the group level. The resulting p-value maps were thresholded at the vertex level with a threshold of $p<0.001$, and corrected for multiple comparisons with a family-wise error rate of $p_{FWE}<0.05$ using FreeSurfer's Monte Carlo simulation-based cluster correction with a vertex-wise cluster-forming threshold of $p<0.001$.

Definition of vertices of interest for characterization

To ensure the reliability of the tuning curves used for characterization, we defined a set of vertices of interest for which the encoding signal was sufficiently strong. One set was defined per type of estimate and per subject, as the vertices with the strongest encoding signal are not the same for probability and confidence, and vary across

subjects. Additionally, the number of vertices was adjusted depending on the availability of vertices with a strong enough signal.

The definition was done in two steps: first at the group level, and then at the subject level. At the group level, a region of interest was defined as the union of parcels from the HCP-MMP1.0 atlas (Glasser et al., 2016) containing the significant clusters found in the group-level statistical maps obtained for the corresponding estimate type (clusters shown on Fig. 3 and 4 and listed in Table 1 and Supplementary Table 1). The vertices of interest for each subject were then defined within the group-level region, using the following steps. The R^2 scores obtained with the encoding model were converted into p-values in each subject and vertex. For each vertex and subject, a p-value was calculated from the R^2 score as the probability of obtaining a value at least as large in a distribution of R^2 values obtained when the fMRI data is replaced with white noise (empirically calculated for each subject with 1,000,000 noise samples). Finally, the vertices with FDR-corrected $p<0.05$ were selected as vertices of interest (controlling the False Discovery Rate using the Benjamini-Hochberg procedure)).

As mentioned in the Results section, we verified the reliability of our estimations within the defined vertices of interest using a test-retest procedure, by fitting the weights of the versatile encoding model on two independent halves of the data and comparing the estimated weights between the two halves.

Characterization of tuning curves

Three characteristic measures were defined and applied to the tuning curves estimated for each vertex of interest.

1) Proportion of tuning curves maximized at non-extreme values. For each tuning curve, we took the input value at which the curve was maximized (i.e. the argmax of the tuning curve function) and treated it as non-extreme if it fell between the lower and upper bounds of the domain with a margin of at least twenty percent relative to the span of the domain. The proportion refers to the proportion of vertices out of all vertices of interest within the subject.

2) Non-monotonicity index. Mathematically speaking, a differentiable function is said to be monotonic if its derivative remains of the same sign over its domain. By extension, we defined the monotonicity index *m(f)* of a function as the absolute value of its average derivative, normalized such that the index of any purely monotonic function is equal to 1. The non-monotonicity index *n(f)* was *1 - m(f)*. It is calculated by the formula $n(f) = 1 - |f(x_{max}) - f(x_{min})| / (f_{max} - f_{min})$, where x_{min} and x_{max} are the lower and upper bounds of the domain, and f_{min} and f_{max} are the minimum and maximum of the function over the domain, respectively. The indices calculated for each tuning curve function were averaged across the vertices of interest to obtain a single value per subject. Note that $m(f)=1$ is a necessary condition for a monotonic function, but not a sufficient one (non-monotonic functions can yield $m(f)=1$).

3) Nonlinearity index. We defined the nonlinearity index at the vertex level as the degree of the polynomial that best fits the estimated tuning curve for that vertex. To avoid favoring larger degrees due to overfitting, we used the tuning curves estimated on two independent halves of the data, one to fit the polynomial model for each degree, the other one to measure the variance explained (R^2) by the fitted polynomials. The nonlinearity index was equal to the degree of the polynomial that led to the best R^2 score. To summarize the computed index values at the subject level, we took the median across vertices. At the group level, we tested statistical significance using a Kruskal-Wallis test by ranks (a nonparametric method for testing whether samples originate from the same distribution).

Decoding of probability and confidence

The same method was used for decoding probability and confidence; here we explain the method in the case of probability. Decoding was performed at the level of each session and subject, and it leveraged the methods used for the encoding analysis. The encoding model characterized a pattern of voxel responses to (overlapping) bins of probability as a set of regression weights assigned to Gaussian basis functions. The decoder aimed to identify the probability bin corresponding to the response pattern measured across several voxels in a test session. Decoding accuracy was assessed with leave-one-session-out cross-validation. The encoding model was fitted (with 5 basis functions instead of 10) separately in the training set (three sessions) and the left-out session. The decoder assigned a probability bin to a given response pattern on the test set by identifying the probability bin eliciting the

most similar response pattern in the training set. More precisely, the decoder used the correlation distance $D(j, k)$ between the response patterns corresponding to the j-th probability bin in the test session and the k-th bin in the training set, and looked for the k that minimized D for a given j. The correlation distance is one minus the Pearson correlation (Walther et al., 2016) qualitatively similar results, although inferior, were obtained with the Euclidean distance.

The decoder was applied to selected voxels in each parcel of the Glasser et al atlas (Glasser et al., 2016). This atlas is available in the fsaverage template used for anatomical normalization (https://figshare.com/articles/HCP-MMP1_0_projected_on_fsaverage/3498446). The atlas was projected onto the native anatomical surface of each subject using the inverse normalization transform, and projected back into their native volume, both with FreeSurfer. The voxels corresponding to each parcel in the functional images were identified based on the coregistration of functional and anatomical images. Decoding was applied to the 100 voxels with the largest $z\text{-}R^2$ value, estimated within the training sessions only.

For each subject, decoding accuracy was assessed for each probability bin as the fraction of correctly assigned probability bins (chance level is 1 out of 5 possibilities) on each left-out session, and averaged across the left-out sessions. Some probability bins concerned fewer than 5% of stimuli in some test sessions, which we deemed unreliable; the corresponding session-level accuracy was omitted from the average across sessions. More precisely, we considered that a stimulus fell in a given bin if its probability elicited more than 10% of the maximum value of the corresponding basis function. The significance of decoding accuracy was assessed, for each parcel, with a two-sided t-test against chance-level accuracy at the group level, and corrected for multiple comparisons across parcels with false discovery rate (0.05) correction (Benjamini-Hochberg procedure).

Analysis of Representational Dissimilarity Matrices (RDMs)

The same method was used for probability and confidence; here we explain the method in the case of probability. We analyzed the distance matrices $D(j, k)$ used for the decoding process, which are also known as representational dissimilarity matrices (RDM). For each subject the RDM of each parcel was the average RDM obtained across cross-validation folds (ignoring the rows j of each RDM corresponding to probability bins concerning fewer than 5% of stimuli in a session). Note that since RDMs were estimated with leave-one-session-out cross-validation, they are not symmetric and their diagonal is not 0. The empirical RDMs were compared to the RDMs of different neural codes, namely, the identity RDM that should arise from a highly nonlinear code, and the graded RDM that should arise from a linear (and more generally monotonic) code. We estimated the regression weights corresponding to each theoretical RDM in a multiple regression analysis; theoretical RDMs were z-scored to make the regression weights commensurable.

Data availability

Behavioral data, raw and preprocessed MRI data will be made available on a public repository upon acceptance.

Code availability

All analysis code to reproduce the reported results and figures from the shared data will be made available on GitHub upon acceptance.

Acknowledgements

This work was supported by funding from the French National Research Agency (ANR grand #18-CE37-0010-01) and from the European Research Council (ERC grant #947105) to F. M. C.F. was supported by a PhD fellowship from ENS Paris-Saclay (France).

Author contributions

CF: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing—original draft, Writing—review & editing, Visualization; TB: Conceptualization, Methodology, Software, Investigation; SD: Methodology, Software; BT: Conceptualization, Methodology, Writing—review & editing; EE: Conceptualization, Methodology, Writing—review & editing; FM: Conceptualization, Methodology, Software, Formal analysis, Resources, Writing—original draft, Writing—review and editing, Visualization, Supervision, Project administration, Funding acquisition.

Competing interests

The authors declare no competing interests.

References

Bach, D. R., Hulme, O., Penny, W. D., & Dolan, R. J. (2011). The Known Unknowns: Neural Representation of Second-Order Uncertainty, and Ambiguity. *The Journal of Neuroscience*, *31*(13), 4811–4820. https://doi.org/10.1523/JNEUROSCI.1452-10.2011

Barretto-García, M., de Hollander, G., Grueschow, M., Polanía, R., Woodford, M., & Ruff, C. C. (2023). Individual risk attitudes arise from noise in neurocognitive magnitude representations. *Nature Human Behaviour*, *7*(9), 1551–1567. https://doi.org/10.1038/s41562-023-01643-4

Bishop, C. M. (2007). *Pattern Recognition and Machine Learning*. Springer.

Bounmy, T., Eger, E., & Meyniel, F. (2023). A characterization of the neural representation of confidence during probabilistic learning. *NeuroImage*, 119849.

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science (New York, N.Y.)*, *340*(6128), 95–98. https://doi.org/10.1126/science.1233912

Dayan, P., & Abbott, L. F. (2005). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (1st ed.). The MIT Press.

Diaconescu, A. O., Mathys, C., Weber, L. A. E., Daunizeau, J., Kasper, L., Lomakina, E. I., Fehr, E., & Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical Bayesian learning. *PLoS Computational Biology*, *10*(9), e1003810. https://doi.org/10.1371/journal.pcbi.1003810

Diehl, G. W., Hon, O. J., Leutgeb, S., & Leutgeb, J. K. (2017). Grid and nongrid cells in medial entorhinal cortex represent spatial location and environmental features with complementary coding schemes. *Neuron*, *94*(1), 83–92.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. https://doi.org/10.1016/j.neuroimage.2007.09.034

Eger, E. (2016). Chapter 1—Neuronal foundations of human numerical representations. In M. Cappelletti & W. Fias (Eds.), *Progress in Brain Research* (Vol. 227, pp. 1–27). Elsevier. https://doi.org/10.1016/bs.pbr.2016.04.015

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering Cortical Number Coding from Human Brain Activity Patterns. *Current Biology*, *19*(19), 1608–1615. https://doi.org/10.1016/j.cub.2009.08.047

Ferrari-Toniolo, S., & Schultz, W. (2023). Reliable population code for subjective economic value from heterogeneous neuronal signals in primate orbitofrontal cortex. *Neuron*. https://doi.org/10.1016/j.neuron.2023.08.009

Findling, C., Hubert, F., Laboratory, I. B., Acerbi, L., Benson, B., Benson, J., Birman, D., Bonacchi, N., Carandini, M., Catarino, J. A., Chapuis, G. A., Churchland, A. K., Dan, Y., DeWitt, E. E., Engel, T. A., Fabbri, M., Faulkner, M., Fiete, I. R., Freitas-Silva, L., … Pouget, A. (2023). *Brain-wide representations of prior information in mouse decision-making* (p. 2023.07.04.547684). bioRxiv. https://doi.org/10.1101/2023.07.04.547684

Fiorillo, C. D. (2003). Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. *Science*, *299*(5614), 1898–1902. https://doi.org/10.1126/science.1077349

Foucault, C., & Meyniel, F. (2023). Two determinants of dynamic adaptive learning for magnitudes and probabilities. *bioRxiv*, 2023–08.

Franke, R. (1982). Scattered data interpolation: Tests of some methods. *Mathematics of Computation*, *38*(157), 181–200.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., & Penny, W. (2007). *Statistical parametric mapping: The analysis of functional brain images* (Academic Press).

Gallistel, C. R. (2017). The Coding Question. *Trends in Cognitive Sciences*, *21*(7), 498–508. https://doi.org/10.1016/j.tics.2017.04.012

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal Population Coding of Movement Direction. *Science*, *233*(4771), 1416–1419. https://doi.org/10.1126/science.3749885

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., & Jenkinson, M. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178.

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Hardcastle, K., Maheswaranathan, N., Ganguli, S., & Giocomo, L. M. (2017). A multiplexed, heterogeneous, and adaptive code for navigation in medial entorhinal cortex. *Neuron*, *94*(2), 375–387.

Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic Representation of Numerosity in the Human Parietal Cortex. *Science*, *341*(6150), 1123–1126. https://doi.org/10.1126/science.1239052

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574–591. https://doi.org/10.1113/jphysiol.1959.sp006308

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Iigaya, K. (2016). Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *eLife*, *5*. https://doi.org/10.7554/eLife.18073

Jacob, S. N., & Nieder, A. (2009). Tuning to non-symbolic proportions in the human frontoparietal cortex. *European Journal of Neuroscience*, *30*(7), 1432–1442. https://doi.org/10.1111/j.1460-9568.2009.06932.x

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience*, *9*(5), 690–696. https://doi.org/10.1038/nn1691

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227–231. https://doi.org/10.1038/nature07200

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology*, *55*(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the Onion of Brain Representations. *Annual Review of Neuroscience*, *42*(1), 407–432. https://doi.org/10.1146/annurev-neuro-080317-061906

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kutlu, M. G., Zachry, J. E., Melugin, P. R., Cajigas, S. A., Chevee, M. F., Kelly, S. J., Kutlu, B., Tian, L., Siciliano, C. A., & Calipari, E. S. (2021). Dopamine release in the nucleus accumbens core signals perceived saliency. *Current Biology*, *31*(21), 4748-4761.e8. https://doi.org/10.1016/j.cub.2021.08.052

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, *18*(8), 1159–1167. https://doi.org/10.1038/nn.4064

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron*, *64*(3), 431–439. https://doi.org/10.1016/j.neuron.2009.09.040

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157. https://doi.org/10.1038/35084005

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438. https://doi.org/10.1038/nn1790

Marshall, T. R., Ruesseler, M., Hunt, L. T., & O'Reilly, J. X. (2022). *The representation of priors and decisions in parietal cortex* (p. 2021.05.03.442155). bioRxiv. https://doi.org/10.1101/2021.05.03.442155

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron*, *84*(4), 870–881. https://doi.org/10.1016/j.neuron.2014.10.013

Meyniel, F., & Dehaene, S. (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proceedings of the National Academy of Sciences*, *114*(19), E3859–E3868.

Monosov, I. E., & Hikosaka, O. (2013). Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. *Nature Neuroscience*, *16*(6), 756–762. https://doi.org/10.1038/nn.3398

Moser, E. I., Kropff, E., & Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annual Review of Neuroscience*, *31*, 69–89. https://doi.org/10.1146/annurev.neuro.31.061307.090723

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410. https://doi.org/10.1016/j.neuroimage.2010.07.073

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An Approximately Bayesian Delta-Rule Model

Explains the Dynamics of Belief Updating in a Changing Environment. *The Journal of Neuroscience*, *30*(37), 12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366–382. https://doi.org/10.1038/nrn.2016.40

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Comput Biol*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

O'Reilly, J. X., Schuffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, *110*(38), E3660–E3669. https://doi.org/10.1073/pnas.1305373110

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience*, *11*(1), 95–102. https://doi.org/10.1038/nn2020

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., & Frith, C. D. (2007). How the brain translates money into force: A neuroimaging study of subliminal motivation. *Science (New York, N.Y.)*, *316*(5826), 904–906. https://doi.org/10.1126/science.1140459

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045. https://doi.org/10.1038/nature05051

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*, *77*(5), 534–540. https://doi.org/10.1001/jamapsychiatry.2019.3671

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci*, *9*(7), 545–556. https://doi.org/10.1038/nrn2357

Sahani, M., & Dayan, P. (2003). Doubly Distributional Population Codes: Simultaneous Representation of Uncertainty and Multiplicity. *Neural Computation*, *15*(10), 2255–2279. https://doi.org/10.1162/089976603322362356

Sara, S. J., & Bouret, S. (2012). Orienting and Reorienting: The Locus Coeruleus Mediates Cognition through Arousal. *Neuron*, *76*(1), 130–141. https://doi.org/10.1016/j.neuron.2012.09.011

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Tomov, M. S., Truong, V. Q., Hundia, R. A., & Gershman, S. J. (2020). Dissociable neural correlates of uncertainty underlie different exploration strategies. *Nature Communications*, *11*(1), 2371. https://doi.org/10.1038/s41467-020-15766-z

van Bergen, R. S., Ma, W. J., Pratte, M. S., & Jehee, J. F. M. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience*, *18*(12), 1728–1730. https://doi.org/10.1038/nn.4150

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, *145, Part B*, 166–179. https://doi.org/10.1016/j.neuroimage.2016.10.038

Walker, E. Y., Pohl, S., Denison, R. N., Barack, D. L., Lee, J., Block, N., Ma, W. J., & Meyniel, F. (2023). Studying the neural representations of uncertainty. *Nature Neuroscience*, 1–11. https://doi.org/10.1038/s41593-023-01444-y

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Wang, L., Amalric, M., Fang, W., Jiang, X., Pallier, C., Figueira, S., Sigman, M., & Dehaene, S. (2019). Representation of spatial sequences using nested rules in human prefrontal cortex. *NeuroImage*, *186*, 245–255. https://doi.org/10.1016/j.neuroimage.2018.10.061

Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148), 1075–1080. https://doi.org/10.1038/nature05852

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic Interpretation of Population Codes. *Neural Computation*, *10*(2), Article 2. https://doi.org/10.1162/089976698300017818

Figures

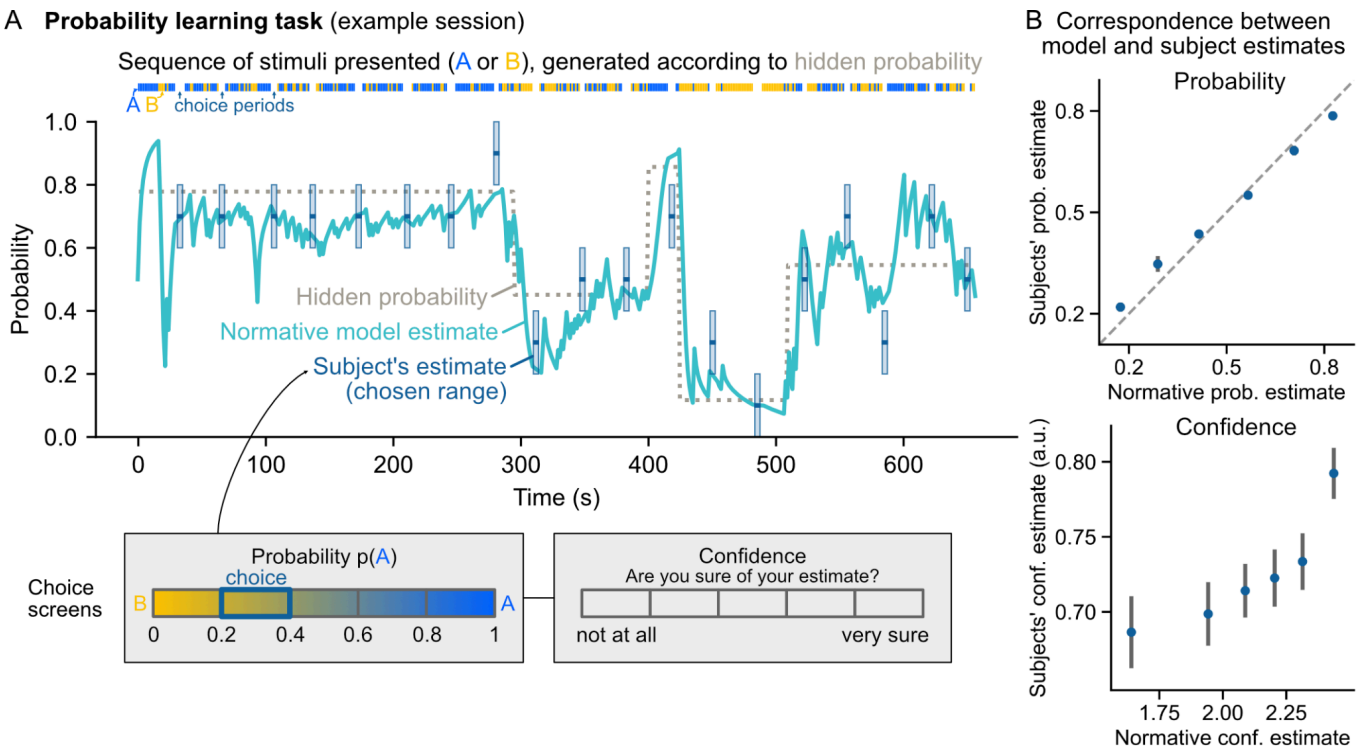
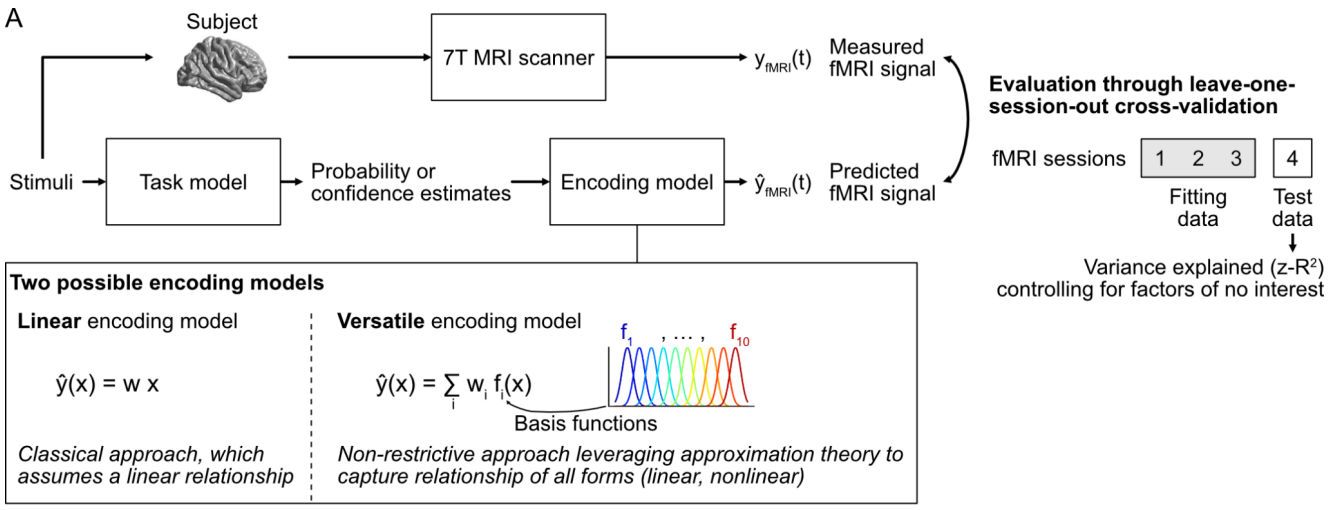
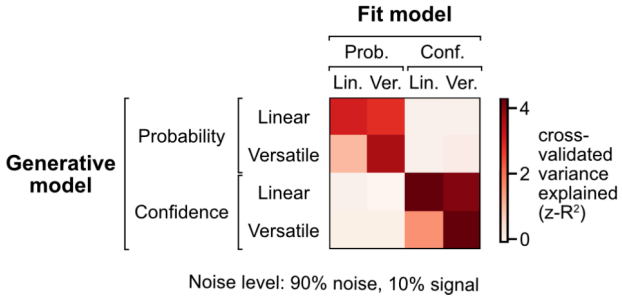


Figure 1. Task and variables of interest. (A) Task session. Visual stimuli of two types, A or B, generated according to a hidden probability p(A), are presented successively to the subject. During the sequence, the hidden probability changes at random and unpredictable times. Throughout the session, the subject must estimate the current value of the hidden probability. Occasionally between two stimuli, a choice period occurs during which the subject chooses a range for their current probability estimate and a confidence level associated with their estimate. The normative model of the task is used to obtain the trial-by-trial estimates that the subject should neurally represent. The values of these different variables are displayed above for an example session. (B) Behavioral results: The estimates chosen by the subjects and the normative estimates are very close for probability (top), and quite correlated for confidence (bottom). Subjects' estimates were binned into six equal quantiles of normative estimate, averaged at the subject level and then at the group level. Subjects' confidence was recorded from 0 to 1; normative confidence is in log precision units (hence the different scales). Dots and error bars show group-level mean±s.e.m. Dashed diagonal is the identity line.



B Validation of the approach through simulation

- Generate noisy fMRI signals encoding probability or confidence according to a given model
- Run the encoding model pipeline on the generated data



Noise level: 90% noise, 10% signal

Figure 2. Modeling the encoding of probability or confidence estimates in voxel-wise fMRI signals. (A) Schematic of the encoding models and their evaluation against fMRI data. For each session, the sequence of stimuli presented to the subject is given to the task model to obtain *x*, the probability or confidence estimates on each trial. These estimates are then given to the encoding model to predict *y*, the fMRI signal time series in a voxel. Two classes of encoding models were tested: the linear class, in which the fMRI signal *y* in a given voxel is a linear function of the estimate *x*, and the versatile class, in which the fMRI signal in a given voxel is a weighted sum of basis functions (f_i) that can approximate linear and nonlinear functions of *x*. To evaluate the models, we used a cross-validation procedure in which three out of four sessions were used to fit the encoding model, and the left-out session was used to measure the predictive accuracy of the model (using the coefficient of determination) after controlling for factors of no interest. (B) Simulation results validating the approach end-to-end. Noisy fMRI data for one experiment were generated assuming a given model of neural activity, and the generated data were used to evaluate each of the encoding models using the procedure described in (A). The matrix shows the average score obtained for each evaluated model and each possible generative model across simulated experiments. Linear models explain the data well only when the generative model is linear, whereas versatile models explain the data well for both classes of generative model. Probability and confidence are well separated by the models.

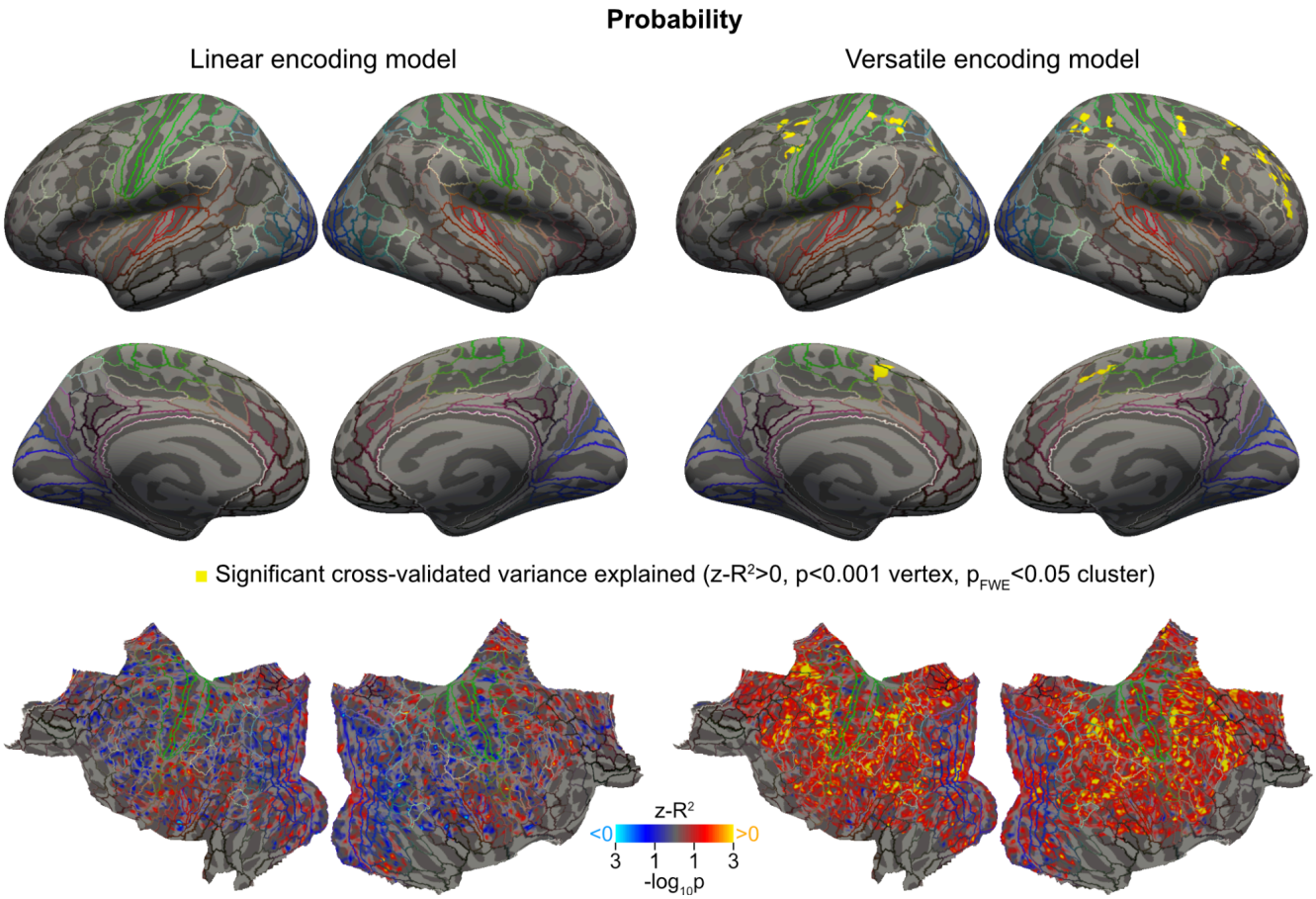


Figure 3. A neural code for probabilities revealed by whole-cortex analysis in prefrontal and parietal cortex, that only the versatile model is able to explain. Cortical maps above show the significance of the cross-validated performance of the linear and versatile encoding models. Top maps show the significant regions after thresholding at p<0.001 and FWE cluster correction at $p_{FWE} < 0.05$. Bottom maps show the p-values without thresholding or correction on a flattened view of the cortex. P-values correspond to the group-level significance of *z-R²* scores obtained across subjects (cold colors for negative scores, hot for positive scores). The HCP-MMP1.0 parcellation is indicated by colored lines (Glasser et al., 2016).

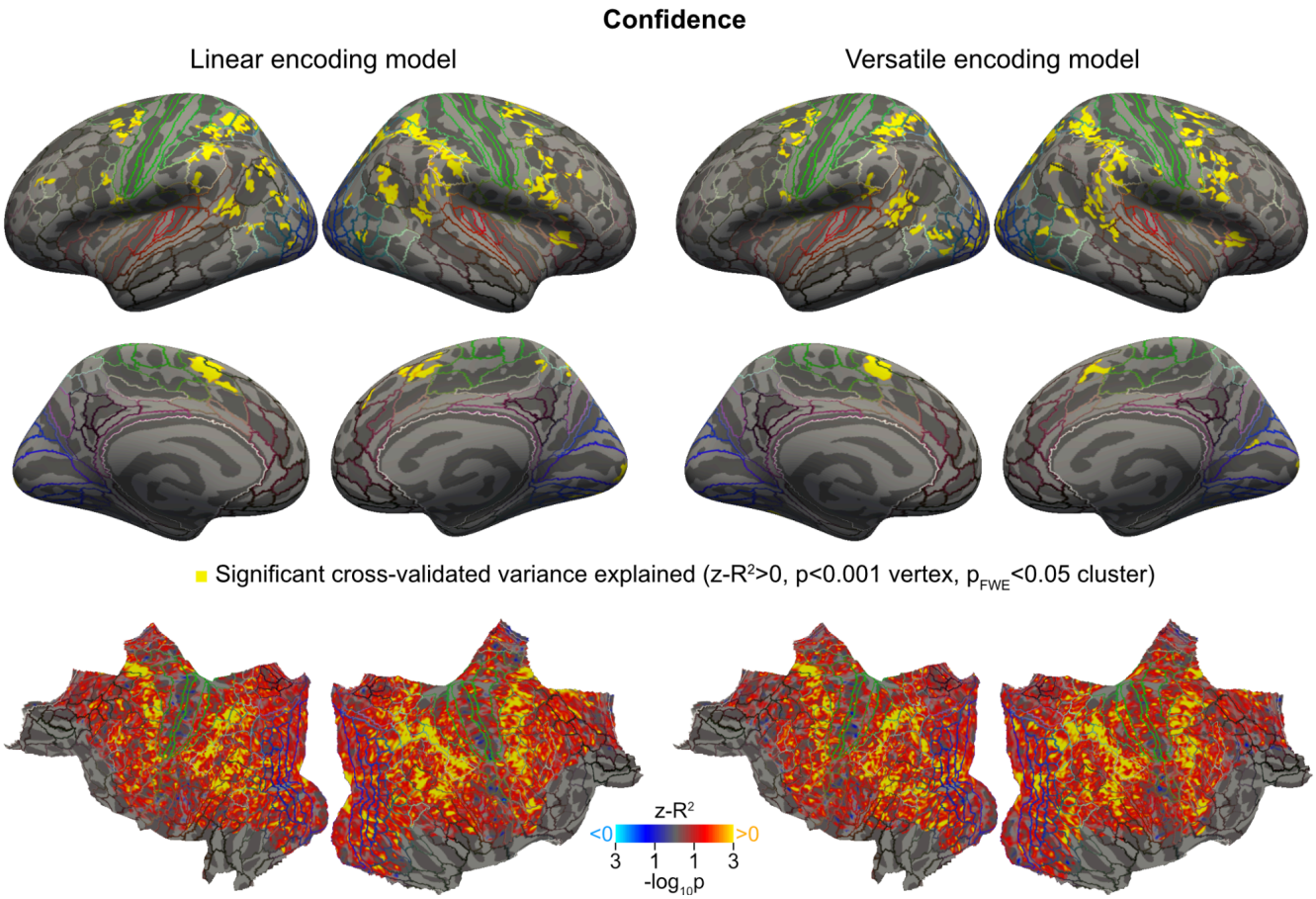


Figure 4. Neural code for confidence. Both the linear and the versatile models explain the neural encoding of confidence in approximately the same regions of the cortex. Plotting conventions are as in Fig. 3.

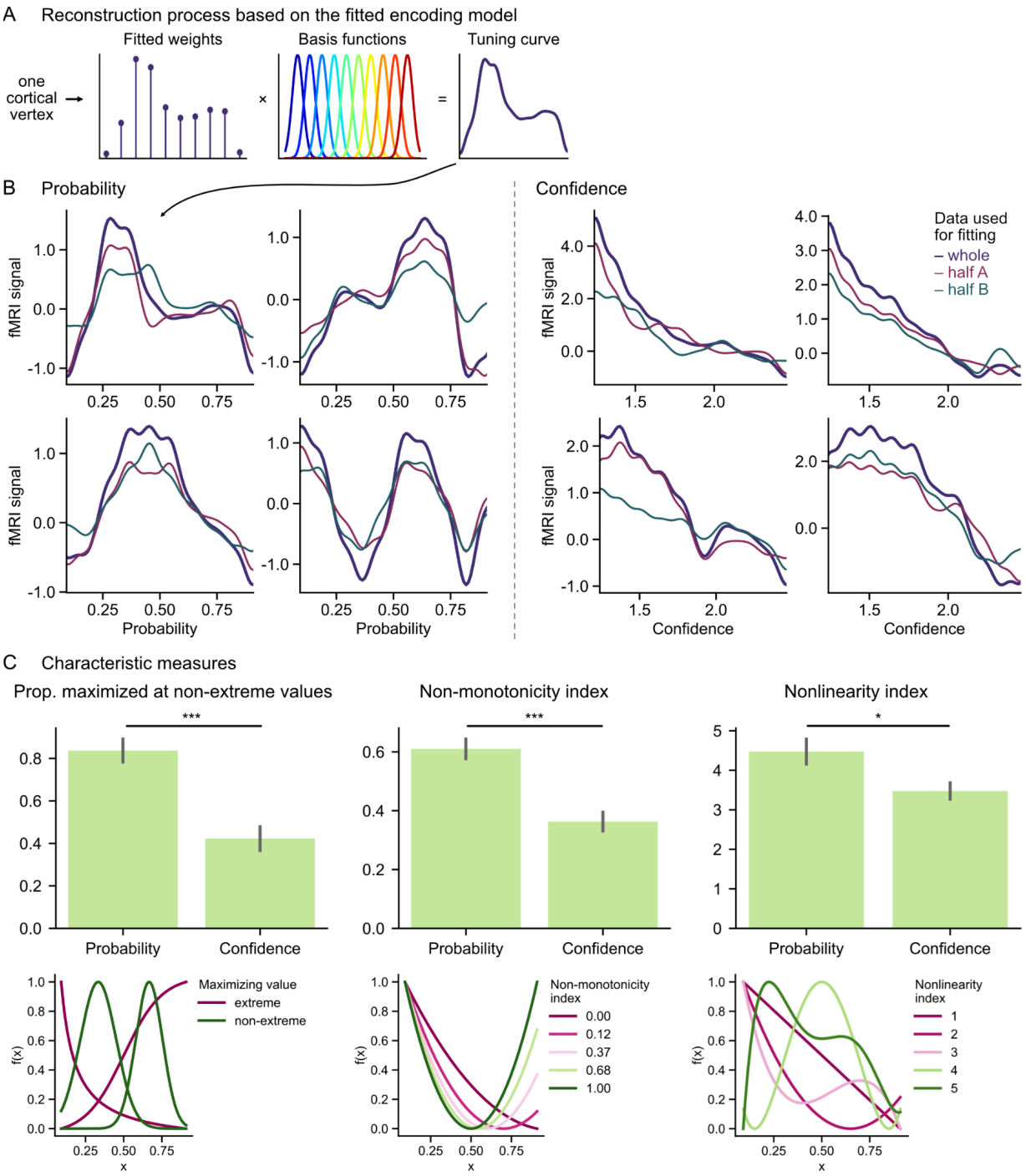


Figure 5. Characterization of the neural encoding of probability and confidence. (A) Schematic of the reconstruction of the tuning curve at the level of a cortical vertex. The weights of the versatile encoding model fitted on the vertex data are used to calculate the tuning curve function, which is equal to the weighted sum of the basis functions. (B) Tuning curves obtained for probability (left) and confidence (right) for example vertices and subjects (out of 30,000 and 120,000 examples, respectively). Within each panel, multiple curves correspond to multiple estimates of the tuning curve for a same vertex: in purple, the curve estimated with the weights fitted on the whole data, red and green, estimated on two independent halves of the data, illustrating test-retest reliability. (C) Quantitative description and comparison of the tuning curves for probability and confidence according to three characteristic measures. The tuning curves for probability are more frequently maximized at non-extreme values, have a higher non-monotonicity index, and a higher nonlinearity index than those for confidence. Bar heights and error bars show mean±s.e.m across subjects. */***: $p<0.05$/$p<0.001$, two-tailed tests. Bottom graphs are illustrations of each of the three measures (*x* denotes the encoded variable, which could be probability or confidence, and *f* a tuning curve function).

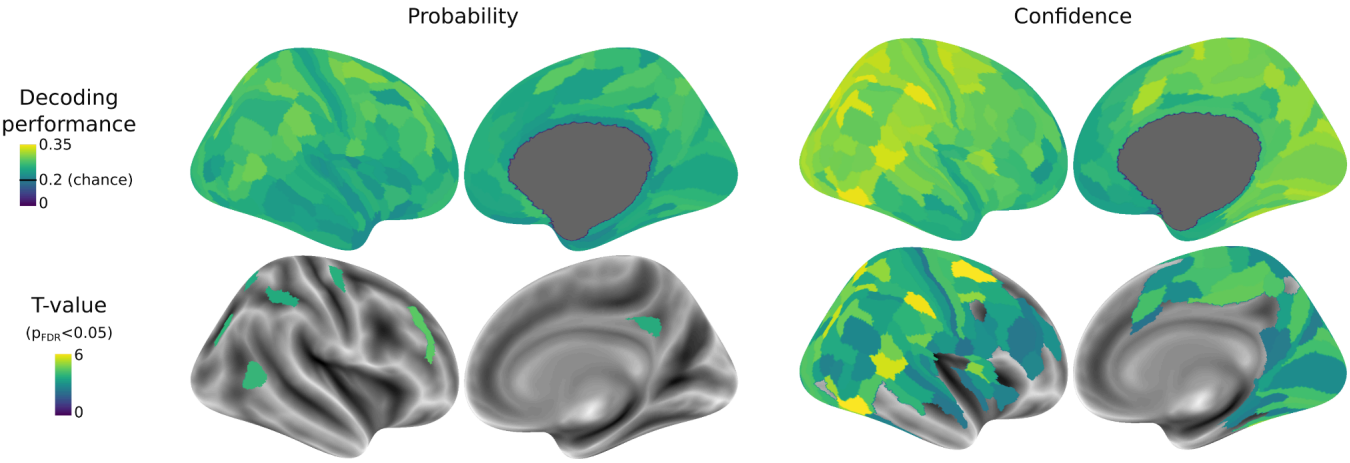


Figure 6. Decoding probability and confidence from the voxel response patterns obtained with the versatile encoding model. Group-level decoding accuracy for each cortical region (180 bihemispheric parcels from the HCP-MMP1.0 atlas, rendered on the right hemisphere for illustration purpose). Top: Mean accuracy across subjects (chance level is one out of five bins, i.e. 0.2). Bottom: T-values of a two-tailed t-test for accuracy different from chance level. Only regions statistically significant after FDR-correction ($p<0.05$) for multiple comparisons across the 180 regions are displayed (also see Table 2 and 3). The decoding method is based on the similarity (correlation distance) of voxel response patterns estimated by the encoding model in a training set and a test set (see Methods).

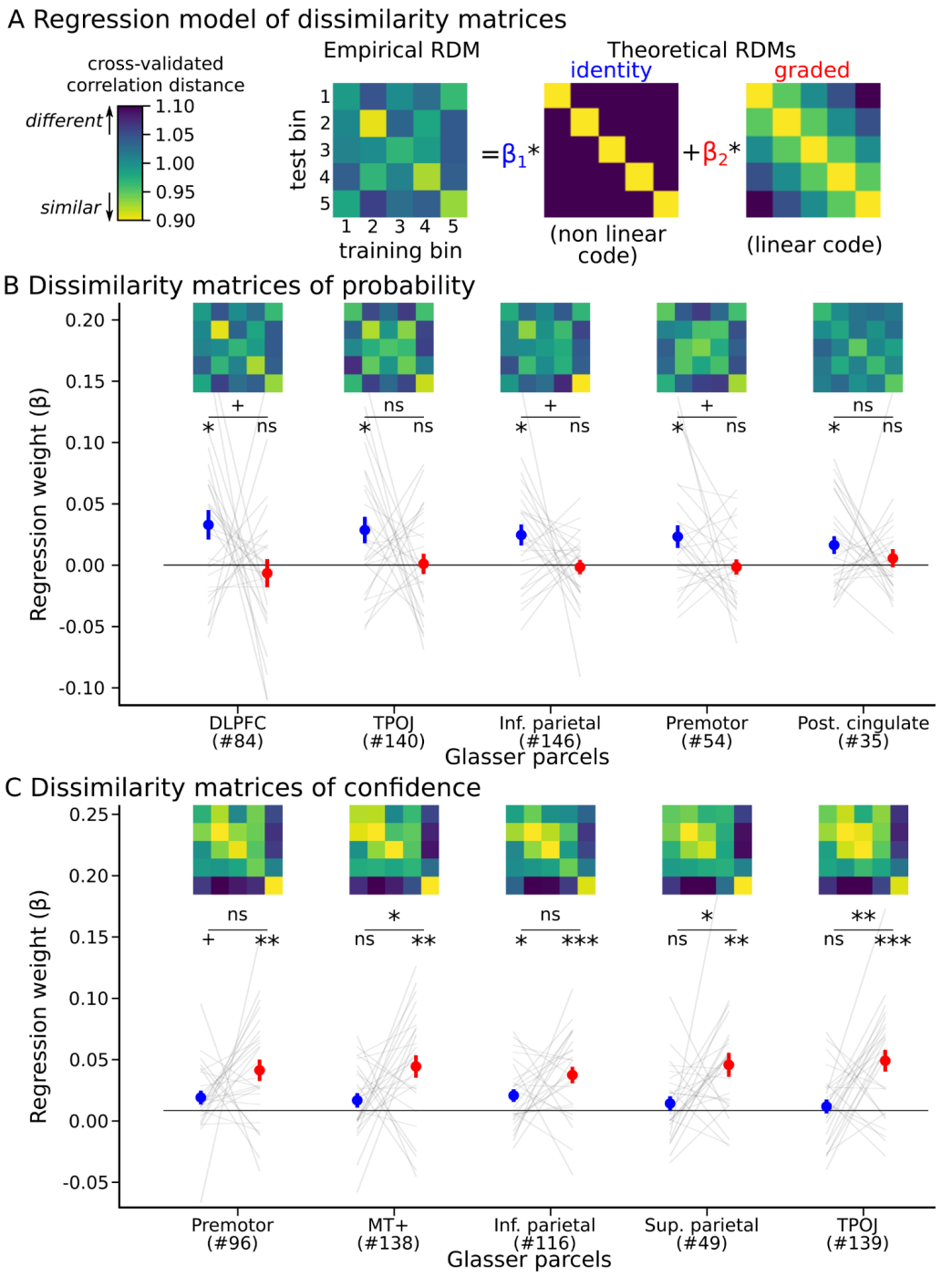


Figure 7: The dissimilarity of voxel response patterns across bins of probability and confidence supports the existence of different types of codes. (A) The representational dissimilarity matrices (RDM) that served for the decoding analysis (Fig. 6) were averaged across test sessions and analyzed with a regression analysis at the subject-level. Two theoretical RDMs were considered in the regression: the "identity" RDM, that should result from a highly nonlinear (and non-monotonic) code, and the "graded" RDM, that should result from a linear (or more generally, monotonic) code. (B) Average RDM of probability levels and regression results in each of the 5 regions with top decoding significance for probability. Each line corresponds to a participant; colored dots and error bars show the average and S.E.M. (C) Same for confidence. In B and C, the same color bar (shown in A) is used for all RDMs. Each set of regression coefficients was tested at the group level against 0 (two-tailed t-test) and compared with a paired t-test (two-tailed); ns: p>0.1; +: p<0.1; *: p<0.05; **:p<0.005; ***:p<0.0005. The parcel numbers correspond to the Glasser parcellation (Glasser et al., 2016); 84: Dorsolateral Prefrontal, Area 46; 140: Temporo-Parieto-Occipital Junction, Area 2, 146: Inferior Parietal, Area 0; 54: Premotor, Dorsal Area 6d, 35: Posterior Cingulate, area 31pv; 96: Premotor, Area 6 anterior; 138: MT+ Complex and Neighboring Visual Areas, Area PH; 116: Inferior Parietal, Area PFt; 49: Superior Parietal, Ventral IntraParietal Complex, 139: Temporo-Parieto-Occipital Junction, Area 1.

Table 1. Significant clusters explained by the versatile encoding model for probability. The names of the cortical areas and parcels refer to the HCP-MMP1.0 atlas (Glasser et al., 2016). Peak x, y, z are MNI coordinates.

Size (mm²)	Num. vertices	Hemisphere	Lobe	Cortical area	Parcel	Peak x	Peak y	Peak z	Peak -log10(p)	Cluster p_{FWE}
160.4	232	Right	Fr	Dorsolateral_Prefrontal	a9-46v_R	38.9	45.2	16	6.4097	0.0002
157.4	183	Left	Occ	Early_Visual	V2_L	-25.1	-97.9	-2.7	4.8187	0.0002
139.01	236	Left	Fr	Paracentral_Lobular_and_Mid_Cingulate	SCEF_L	-9.5	5.3	55.9	4.8964	0.0002
108.66	227	Right	Fr	Premotor	FEF_R	42.2	-5.8	50.7	6.6067	0.0002
106.8	274	Right	Par	Superior_Parietal	LIPv_R	27.7	-52.8	47.4	5.211	0.0002
101.02	198	Left	Par	Superior_Parietal	7PC_L	-37.4	-50.2	55.6	4.5741	0.0002
100.41	202	Right	Fr	Dorsolateral_Prefrontal	46_R	27.3	29.6	36.2	4.0095	0.0002
98.83	176	Left	Fr	Premotor	6a_L	-29.1	0.6	48.6	5.3333	0.0002
80.1	169	Right	Par	Superior_Parietal	7PC_R	35.1	-49.1	60.5	4.1947	0.0006
71.32	161	Left	Par	Inferior_Parietal	IP1_L	-31.3	-64.2	40.1	5.3758	0.0014
70.44	127	Left	Fr	Premotor	FEF_L	-37.3	-7.6	44.5	4.6641	0.0014
69.62	143	Right	Fr	Paracentral_Lobular_and_Mid_Cingulate	24dd_R	9.5	1.1	53.9	4.8056	0.0008
69.09	132	Left	Fr	Dorsolateral_Prefrontal	8Ad_L	-25.1	24.6	33.7	5.6881	0.0014
67.68	105	Right	Fr	Dorsolateral_Prefrontal	s6-8_R	21.7	20	53.1	4.3053	0.0008
64.72	136	Left	Fr	Premotor	PEF_L	-51.2	-2.6	40.4	4.6327	0.002
55.09	83	Right	Fr	Dorsolateral_Prefrontal	9p_R	19.4	38.9	37	5.5244	0.00599
53.88	95	Right	Fr	Dorsolateral_Prefrontal	46_R	35.1	33.9	27.9	4.6328	0.00719
51.51	93	Right	Fr	Dorsolateral_Prefrontal	9-46d_R	30.3	38.8	20.9	4.7414	0.00918
49.07	57	Left	Occ	Primary_Visual	V1_L	-17.4	-100.9	0.2	4.189	0.01296
48.56	70	Left	Fr	Dorsolateral_Prefrontal	46_L	-39.2	31.5	29.3	4.2904	0.01355
48.14	103	Left	Par	Somatosensory_and_Motor	2_L	-36.9	-35.2	56.7	4.5609	0.01415
45.65	96	Right	Fr	Dorsolateral_Prefrontal	8C_R	33.9	10.8	33.4	4.4846	0.0203
44.86	69	Right	Fr	Dorsolateral_Prefrontal	9-46d_R	25.4	38.8	32.1	4.2132	0.02247
43.94	88	Left	Fr	Premotor	55b_L	-45.1	1.5	46.8	3.7926	0.02366
43.3	56	Right	Fr	Dorsolateral_Prefrontal	9-46d_R	25.2	45.1	29.1	4.2352	0.02702
43.14	84	Right	Par	Inferior_Parietal	IP2_R	50.1	-37	45.4	5.2945	0.02721
41.56	123	Right	Par	Somatosensory_and_Motor	2_R	31.5	-35.1	46.2	3.9812	0.03292
41.06	91	Left	Par	Superior_Parietal	AIP_L	-31.1	-48.1	44.3	4.2868	0.0345
38.14	72	Left	Par	Temporo-Parieto-Occipital_Junction	STV_L	-53.8	-46.6	11.7	4.2271	0.04879

Table 2. Decoding accuracy for probability. Names of cortical areas are as in Table 1. The decoding accuracy is compared to chance level (0.2) with a two-sided t-test. Only parcels with significant decoding accuracy (p_{FDR}<0.05) are reported.

Lobe	Cortical area	Parcel	Mean	T value	P value	p_{FDR}
Fr	Dorsolateral_Prefrontal	46	0.27	4.46	0.0002	0.0327
Par	Temporo-Parieto-Occipital_Junction	TPOJ2	0.27	4.03	0.0006	0.0327
Occ	Inferior_Parietal	IP0	0.27	3.93	0.0007	0.0327
Fr	Premotor	6d	0.28	3.84	0.0009	0.0327
Par	Posterior_Cingulate	31pv	0.27	3.76	0.0011	0.0327
Par	Superior_Parietal	AIP	0.28	3.74	0.0012	0.0327
Par	Superior_Parietal	VIP	0.28	3.70	0.0013	0.0327

Table 3. Decoding accuracy for confidence. Names of cortical areas are as in Table 1. The decoding accuracy is compared to chance level (0.2) with a two-sided t-test. Only parcels with significant decoding accuracy (p_{FDR}<0.05) are reported.

Lobe	Cortical area	Parcel	Mean	T value	P value	p_{FDR}
Fr	Premotor	6a	0.30	6.29	0.0000	0.0002
Temp	MT+_Complex_and_Neighboring_Visual_Areas	PH	0.34	6.18	0.0000	0.0002
Par	Inferior_Parietal	PFt	0.34	5.87	0.0000	0.0003
Par	Superior_Parietal	VIP	0.32	5.72	0.0000	0.0003
Temp	Temporo-Parieto-Occipital_Junction	TPOJ1	0.33	5.59	0.0000	0.0004
Occ	Inferior_Parietal	IP0	0.32	5.52	0.0000	0.0004
Par	Superior_Parietal	LIPd	0.34	5.28	0.0000	0.0006
Occ	MT+_Complex_and_Neighboring_Visual_Areas	V4t	0.31	5.15	0.0000	0.0007
Par	Superior_Parietal	7PC	0.32	5.06	0.0000	0.0007
Temp	Ventral_Stream_Visual	FFC	0.31	5.04	0.0000	0.0007
Fr	Premotor	55b	0.31	5.02	0.0000	0.0007
Ins	Insular_and_Frontal_Opercular	FOP3	0.29	4.99	0.0000	0.0007
Par	Superior_Parietal	MIP	0.32	4.81	0.0001	0.0010
Par	Temporo-Parieto-Occipital_Junction	STV	0.31	4.75	0.0001	0.0010
Occ	Dorsal_Stream_Visual	V7	0.30	4.75	0.0001	0.0010
Par	Early_Auditory	PFcm	0.30	4.75	0.0001	0.0010
Occ	MT+_Complex_and_Neighboring_Visual_Areas	V3CD	0.30	4.66	0.0001	0.0012
Par	Inferior_Parietal	PGs	0.33	4.64	0.0001	0.0012
Par	Paracentral_Lobular_and_Mid_Cingulate	23c	0.28	4.59	0.0001	0.0013
Par	Paracentral_Lobular_and_Mid_Cingulate	5mv	0.28	4.52	0.0002	0.0014
Par	Dorsal_Stream_Visual	IPS1	0.33	4.52	0.0002	0.0014
Occ	Early_Visual	V3	0.31	4.42	0.0002	0.0016
Occ	Posterior_Cingulate	POS2	0.29	4.40	0.0002	0.0016
Temp	Lateral_Temporal	PHT	0.30	4.40	0.0002	0.0016
Par	Posterior_Cingulate	PCV	0.30	4.35	0.0002	0.0018
Fr	Paracentral_Lobular_and_Mid_Cingulate	6mp	0.31	4.34	0.0003	0.0018
Occ	Early_Visual	V4	0.30	4.32	0.0003	0.0018
Fr	Somatosensory_and_Motor	4	0.28	4.31	0.0003	0.0018
Fr	Anterior_Cingulate_and_Medial_Prefrontal	p32pr	0.31	4.29	0.0003	0.0018
Par	Inferior_Parietal	PFop	0.29	4.28	0.0003	0.0018
Fr	Paracentral_Lobular_and_Mid_Cingulate	24dd	0.28	4.26	0.0003	0.0018
Fr	Premotor	6d	0.32	4.21	0.0004	0.0019
Occ	Ventral_Stream_Visual	VMV2	0.28	4.21	0.0004	0.0019
Occ	Early_Visual	V2	0.30	4.20	0.0004	0.0019
Fr	Posterior_Opercular	FOP1	0.27	4.18	0.0004	0.0020
Temp	Ventral_Stream_Visual	VVC	0.30	4.12	0.0004	0.0022
Occ	Ventral_Stream_Visual	PIT	0.29	4.07	0.0005	0.0024
Par	Superior_Parietal	AIP	0.30	4.00	0.0006	0.0028
Temp	Auditory_Association	STSvp	0.28	3.96	0.0007	0.0031
Par	Inferior_Parietal	PF	0.29	3.89	0.0008	0.0035
Fr	Premotor	FEF	0.30	3.89	0.0008	0.0035
Par	Early_Auditory	RI	0.30	3.88	0.0008	0.0035
Par	Somatosensory_and_Motor	2	0.28	3.83	0.0009	0.0038
Fr	Paracentral_Lobular_and_Mid_Cingulate	6ma	0.28	3.78	0.0011	0.0043
Par	Paracentral_Lobular_and_Mid_Cingulate	5m	0.29	3.77	0.0011	0.0043
Fr	Dorsolateral_Prefrontal	8Av	0.27	3.75	0.0011	0.0044
Fr	Insular_and_Frontal_Opercular	FOP4	0.30	3.74	0.0012	0.0044
Par	Temporo-Parieto_Occipital_Junction	PSL	0.30	3.73	0.0012	0.0045
Fr	Dorsolateral_Prefrontal	i6-8	0.27	3.70	0.0013	0.0047
Occ	Inferior_Parietal	PGp	0.28	3.69	0.0013	0.0048
Par	Superior_Parietal	LIPv	0.30	3.67	0.0014	0.0048
Occ	Temporo-Parieto-Occipital_Junction	TPOJ3	0.29	3.67	0.0014	0.0048
Par	Superior_Parietal	7AL	0.31	3.66	0.0014	0.0049
Par	Somatosensory_and_Motor	1	0.29	3.64	0.0015	0.0050
Par	Superior_Parietal	7PI	0.29	3.62	0.0016	0.0051
Par	Inferior_Parietal	IP1	0.29	3.55	0.0018	0.0058

Occ	Dorsal_Stream_Visual	V3A	0.28	3.55	0.0018	0.0058
Temp	Auditory_Association	STSdp	0.29	3.53	0.0020	0.0060
Occ	MT+_Complex_and_Neighboring_Visual_Areas	FST	0.29	3.51	0.0020	0.0060
Par	Inferior_Parietal	PFm	0.27	3.51	0.0020	0.0060
Temp	Early_Auditory	PBelt	0.28	3.51	0.0020	0.0060
Par	Posterior_Cingulate	7m	0.29	3.44	0.0024	0.0070
Occ	MT+_Complex_and_Neighboring_Visual_Areas	MST	0.27	3.44	0.0024	0.0070
Par	Posterior_Cingulate	ProS	0.27	3.42	0.0026	0.0072
Fr	Dorsolateral_Prefrontal	8C	0.27	3.41	0.0026	0.0073
Occ	MT+_Complex_and_Neighboring_Visual_Areas	LO1	0.28	3.39	0.0027	0.0074
Fr	Inferior_Frontal	IFJa	0.27	3.39	0.0027	0.0074
Temp	Early_Auditory	A1	0.27	3.35	0.0030	0.0079
Fr	Premotor	6v	0.27	3.32	0.0032	0.0084
Occ	Posterior_Cingulate	DVT	0.30	3.32	0.0033	0.0084
Temp	Auditory_Association	A4	0.27	3.25	0.0038	0.0097
Ins	Insular_and_Frontal_Opercular	MI	0.26	3.25	0.0039	0.0097
Par	Somatosensory_and_Motor	3b	0.28	3.21	0.0042	0.0103
Occ	Dorsal_Stream_Visual	V3B	0.27	3.20	0.0043	0.0106
Par	Posterior_Opercular	OP4	0.27	3.19	0.0044	0.0107
Par	Inferior_Parietal	IP2	0.26	3.13	0.0051	0.0121
Fr	Dorsolateral_Prefrontal	9-46d	0.27	3.08	0.0058	0.0136
Fr	Paracentral_Lobular_and_Mid_Cingulate	SCEF	0.28	3.05	0.0061	0.0141
Fr	Premotor	6r	0.27	3.01	0.0068	0.0155
Par	Paracentral_Lobular_and_Mid_Cingulate	5L	0.27	2.99	0.0072	0.0160
Occ	Primary_Visual	V1	0.28	2.98	0.0072	0.0160
Par	Superior_Parietal	7Am	0.27	2.95	0.0078	0.0169
Fr	Paracentral_Lobular_and_Mid_Cingulate	24dv	0.26	2.95	0.0078	0.0169
Occ	Ventral_Stream_Visual	V8	0.27	2.93	0.0081	0.0173
Fr	Inferior_Frontal	IFJp	0.28	2.93	0.0082	0.0174
Par	Temporo-Parieto-Occipital_Junction	TPOJ2	0.28	2.90	0.0088	0.0185
Temp	Medial_Temporal	PHA1	0.27	2.87	0.0093	0.0193
Par	Inferior_Parietal	PGi	0.27	2.86	0.0096	0.0196
Temp	Early_Auditory	LBelt	0.27	2.85	0.0098	0.0198
Occ	MT+_Complex_and_Neighboring_Visual_Areas	MT	0.28	2.81	0.0107	0.0215
Par	Posterior_Opercular	OP1	0.25	2.79	0.0114	0.0223
Temp	Medial_Temporal	PreS	0.27	2.78	0.0115	0.0223
Fr	Inferior_Frontal	IFSa	0.27	2.78	0.0115	0.0223
Temp	Lateral_Temporal	TE2p	0.26	2.75	0.0124	0.0237
Fr	Insular_and_Frontal_Opercular	AVI	0.26	2.74	0.0127	0.0240
Occ	MT+_Complex_and_Neighboring_Visual_Areas	LO3	0.28	2.71	0.0135	0.0254
Occ	Dorsal_Stream_Visual	V6A	0.27	2.69	0.0142	0.0264
Temp	Medial_Temporal	PHA2	0.27	2.62	0.0164	0.0301
Occ	Ventral_Stream_Visual	VMV3	0.26	2.61	0.0168	0.0301
Fr	Dorsolateral_Prefrontal	46	0.26	2.61	0.0168	0.0301
Temp	Auditory_Association	A5	0.27	2.61	0.0169	0.0301
Occ	Ventral_Stream_Visual	VMV1	0.27	2.56	0.0190	0.0335
Occ	Dorsal_Stream_Visual	V6	0.26	2.53	0.0202	0.0352
Par	Posterior_Cingulate	POS1	0.26	2.51	0.0210	0.0363
Temp	Auditory_Association	STSda	0.26	2.46	0.0237	0.0405
Fr	Inferior_Frontal	IFSp	0.25	2.45	0.0238	0.0405
Ins	Insular_and_Frontal_Opercular	PoI2	0.26	2.45	0.0241	0.0405
Fr	Somatosensory_and_Motor	3a	0.25	2.43	0.0250	0.0417
Par	Posterior_Cingulate	v23ab	0.26	2.40	0.0266	0.0439
Fr	Dorsolateral_Prefrontal	p9-46v	0.25	2.40	0.0269	0.0441
Fr	Posterior_Opercular	43	0.25	2.37	0.0284	0.0461
Fr	Anterior_Cingulate_and_Medial_Prefrontal	a24pr	0.26	2.34	0.0304	0.0489

Supplementary information

Supplementary Notes

Supplementary Note 1

Equation to be demonstrated:

$$p(h_{t+1} \mid s_{1:t+1}) \propto p(s_{t+1} \mid h_{t+1}) \int p(h_{t+1} \mid h_t)\, p(h_t \mid s_{1:t})\, dh_t \quad [1]$$

Below is the mathematical proof of the formula given by equation [1].

$$p(h_{t+1} \mid s_{1:t+1}) = p(h_{t+1} \mid s_{t+1}, s_{1:t})$$

$$p(h_{t+1} \mid s_{1:t+1}) = p(h_{t+1}, s_{t+1}, s_{1:t}) \,/\, p(s_{t+1} \mid s_{1:t})$$

$$p(h_{t+1} \mid s_{1:t+1}) = p(h_{t+1}, s_{t+1} \mid s_{1:t})\, p(s_{1:t}) \,/\, p(s_{t+1} \mid s_{1:t})$$

$$p(h_{t+1} \mid s_{1:t+1}) = p(s_{t+1} \mid h_{t+1}, s_{1:t})\, p(h_{t+1} \mid s_{1:t})\, p(s_{1:t}) \,/\, p(s_{t+1} \mid s_{1:t})$$

Conditional independence property (1): s_{t+1} is conditionally independent of $s_{1:t}$ given h_{t+1}, therefore:

$$p(h_{t+1} \mid s_{1:t+1}) = p(s_{t+1} \mid h_{t+1})\, p(h_{t+1} \mid s_{1:t})\, p(s_{1:t}) \,/\, p(s_{t+1} \mid s_{1:t}) \quad [2]$$

The first term of equation [2] is the same as that equation [1]. Let's expand the second term of equation [2] using the sum rule.

$$p(h_{t+1} \mid s_{1:t}) = \int p(h_{t+1}, h_t \mid s_{1:t})\, dh_t$$

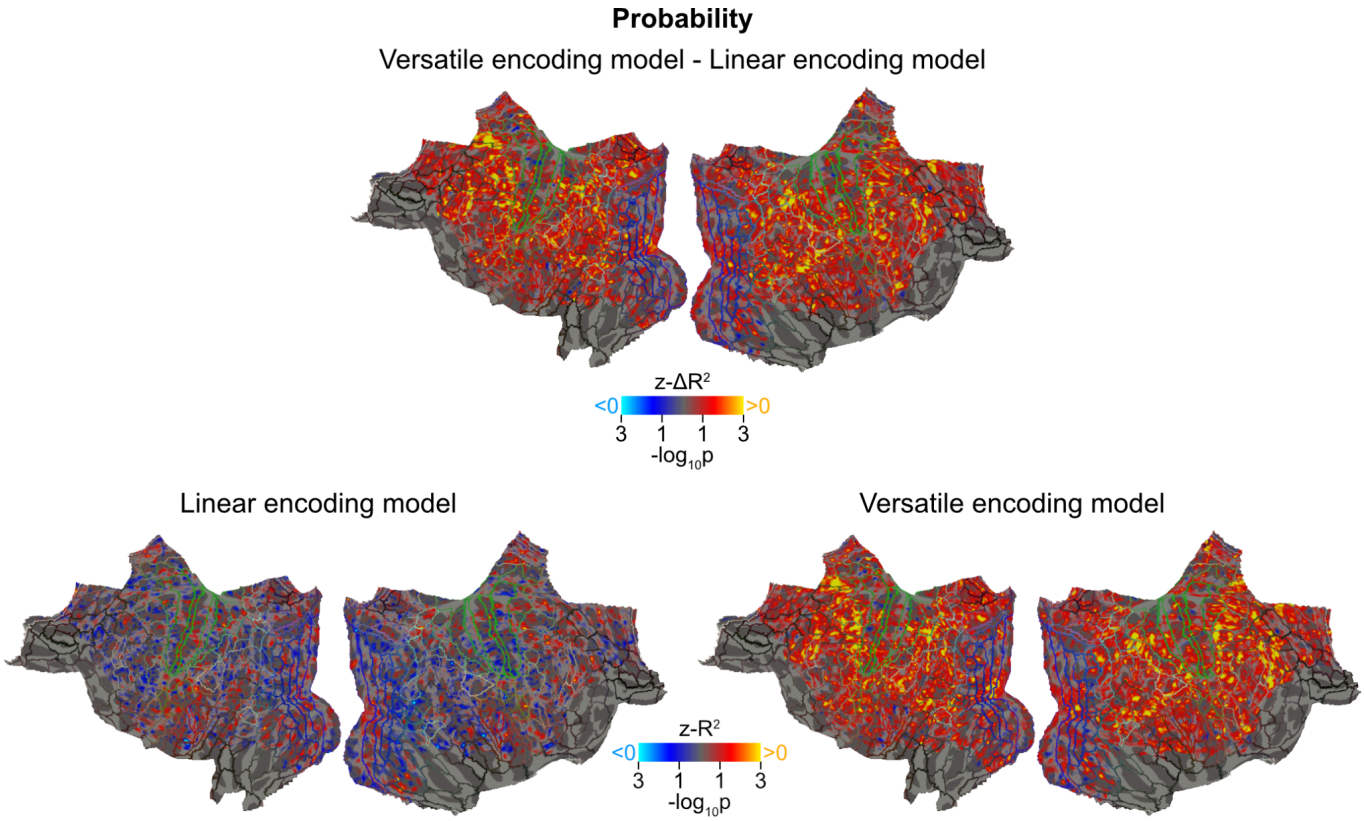
$$p(h_{t+1} \mid s_{1:t}) = \int p(h_{t+1} \mid h_t, s_{1:t})\, p(h_t \mid s_{1:t})\, dh_t$$

Conditional independence property (2) : h_{t+1} is conditionally independent of $s_{1:t}$ given h_t, therefore

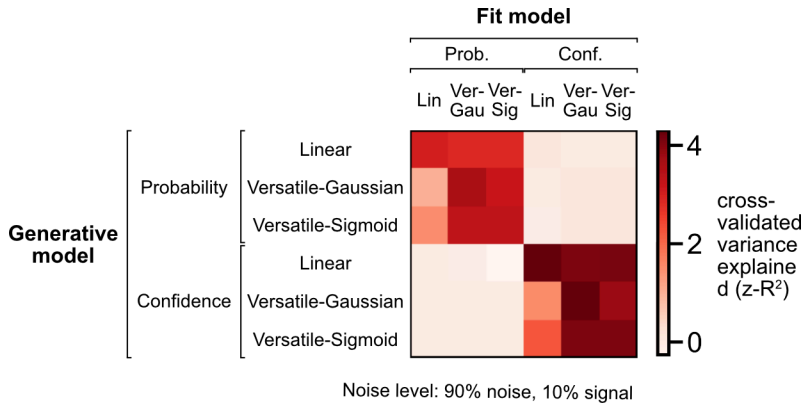
$$p(h_{t+1} \mid s_{1:t}) = \int p(h_{t+1} \mid h_t)\, p(h_t \mid s_{1:t})\, dh_t \quad [3]$$

By combining equations [2] and [3] and omitting the normalization factor $p(s_{1:t}) \,/\, p(s_{t+1} \mid s_{1:t})$, we obtain equation [1] which was to be demonstrated.

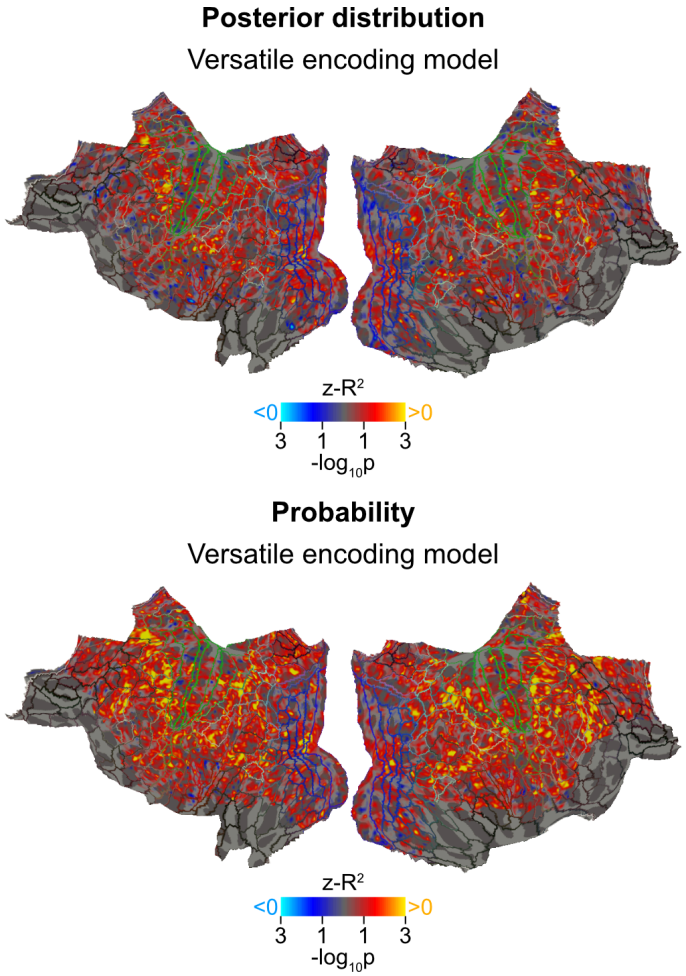
Supplementary Figures



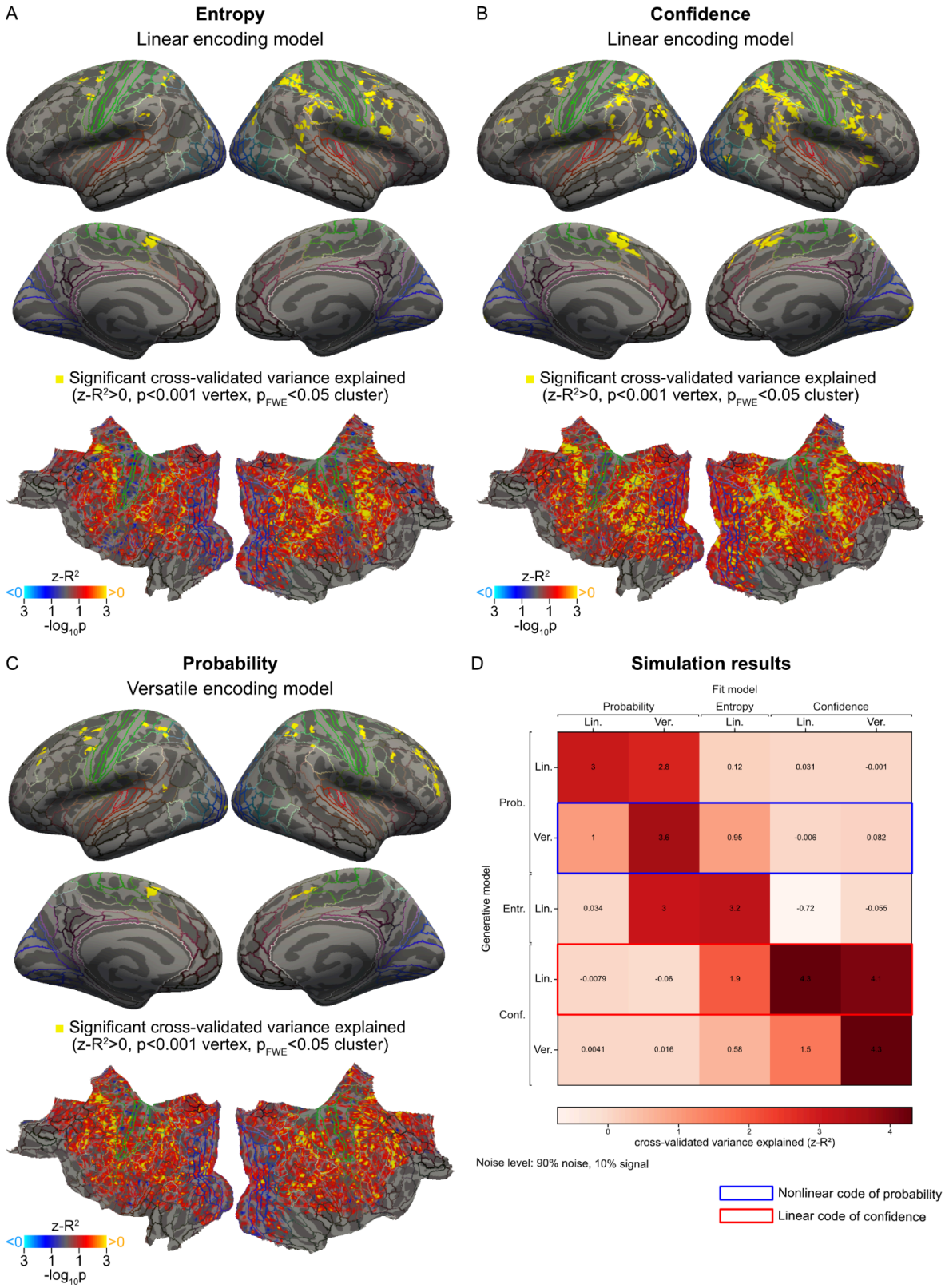
Supplementary Figure 1. Difference in cross-validated variance explained by the versatile and the linear encoding models for probability. Vertex-wise p-values are shown on flattened maps of the cortex. Top: P-values correspond to the group-level significance of z-ΔR^2 scores obtained across subjects (ΔR^2 = R^2[versatile] - R^2[linear], cold/hot colors favor the linear/versatile encoding model respectively, p values are unthresholded and uncorrected). Bottom: Maps of the individual models, repeated from Fig. 3 for illustration purposes. Delineated by colored lines is the HCP-MMP1.0 parcellation (Glasser et al., 2016).



Supplementary Figure 2. Simulation results with Gaussian and sigmoid basis functions. Simulation results obtained as in Fig. 2B after splitting the versatile encoding model into two: one with Gaussian basis functions (the one used in the main text, referred to as Versatile-Gaussian above) and one with sigmoid basis functions (referred to as Versatile-Sigmoid above). The sigmoid basis functions of the Versatile-Sigmoid model are expressed $f_i(x) = 1 / [1 + exp[-k(x-\mu_i)]]$. For comparison, we took the same number of basis functions (10) and the same centers μ_i as for Versatile-Gaussian, and an equivalent slope k was computed from the Versatile-Gaussian's σ using the formula $k = 4 / [(2\pi)^{1/2}\sigma]$.



Supplementary Figure 3. Variance explained by the model encoding the posterior distribution (top) vs. the probability estimate (bottom). Cortical maps as in Fig. 3. See Methods for the encoding model of the posterior distribution. Bottom map is repeated from Fig. 3 for illustration purposes.



Supplementary Figure 4. The (linear) effect of confidence and the (nonlinear) effect of probability are not confounded by a (linear) effect of entropy. (A) Cortical maps for the linear model of entropy, which quantifies the extent to which the next stimulus is unpredictable (entropy is maximum for p(A)=0.5 and gradually decreases as p(A) deviates from 0.5). Plotting conventions are as in Fig. 3. Panels B and C for confidence and probability are reproduced from Fig. 4 and Fig. 3, respectively, to facilitate comparison with A. Note that the regression models in B and C include entropy as an additional regressor, but that confidence and probability are not included as additional regressors in A. The effect of entropy seen in A is less widespread than

the effect of confidence, and largely overlaps with the latter; the effect seen in A is thus likely to arise from the small correlation that exists between entropy and confidence. The effect of confidence we observed in B cannot be reduced to an effect of entropy because the latter effect is included in the regression model, and the effect of confidence is more widespread than the effect of entropy. The (nonlinear) effect of probability in C is also not confounded by a (linear) effect of entropy since entropy is included as an additional regressor, and since the effects of probably and entropy are anatomically distinct (at least in the anterior part of the dorsolateral prefrontal cortex). (D) Simulation results. Plotting conventions are as in Fig. 1B. Note that a versatile model of probability explains the data generated by a linear model of entropy very well, but the converse is not true. Also note that a linear model of entropy can only partially explain the data generated by a linear model of confidence. The pattern of results observed in the data (panels A-C) are consistent with the results obtained by simulations.

Supplementary Table 1. Significant clusters explained by the linear encoding model for confidence. The names of the cortical areas and parcels refer to the HCP-MMP1.0 atlas (Glasser et al., 2016). Peak x, y, z are MNI coordinates.

Size (mm²)	Num. vertices	Hemisphere	Lobe	Cortical area	Parcel	Peak x	Peak y	Peak z	Peak -log10(p)	Cluster p_{FWE}
1562.62	3701	Right	Par	Inferior_Parietal	IP2_R	49.2	-37.8	43	5.9237	0.0002
369.56	695	Left	Fr	Paracentral_Lobular_and_Mid_Cingulate	SCEF_L	-9.4	4.5	56.4	5.7647	0.0002
341.32	676	Right	Par	Inferior_Parietal	PGi_R	47.1	-63.1	31.4	5.5385	0.0002
335.91	791	Left	Par	Superior_Parietal	7PC_L	-36.9	-50.4	54.9	6.1673	0.0002
257.7	601	Left	Par	Inferior_Parietal	PF_L	-49	-38.4	43.1	6.115	0.0002
202.25	328	Right	Fr	Premotor	6r_R	45.8	10.1	17.9	4.4931	0.0002
202.14	389	Left	Fr	Premotor	FEF_L	-37	-8.2	48.3	5.2117	0.0002
198.63	465	Right	Fr	Premotor	6a_R	33.8	-11	55.5	4.7601	0.0002
195.4	402	Right	Fr	Paracentral_Lobular_and_Mid_Cingulate	SCEF_R	7.7	5.8	51.2	4.6558	0.0002
193.77	432	Right	Fr	Premotor	6v_R	56.1	6.7	25.7	5.6508	0.0002
189.66	567	Right	Temp	Temporo-Parieto-Occipital_Junction	TPOJ1_R	44.9	-47.4	12.2	5.1703	0.0002
179.33	371	Left	Par	Temporo-Parieto-Occipital_Junction	STV_L	-55.6	-48.2	13.3	6.1347	0.0002
175.3	458	Left	Par	Superior_Parietal	AIP_L	-30.9	-38.3	41.6	5.2929	0.0002
171.23	209	Right	Occ	Primary_Visual	V1_R	17.8	-94.6	0	4.1598	0.0002
164.69	241	Right	Fr	Premotor	6a_R	30.6	1.4	49.2	4.9644	0.0002
155.42	293	Left	Fr	Premotor	6a_L	-29	-2.3	45.8	5.92	0.0002
143.36	293	Right	Fr	Paracentral_Lobular_and_Mid_Cingulate	6ma_R	16.3	-0.9	66.3	4.8137	0.0002
142.98	326	Right	Par	Inferior_Parietal	PF_R	48.4	-37.2	26.8	4.2616	0.0002
137	362	Right	Ins	Insular_and_Frontal_Opercular	FOP5_R	32.5	27.2	7.9	6.2325	0.0002
134.03	253	Left	Par	Inferior_Parietal	PGs_L	-40.5	-69.8	32.7	6.4739	0.0002
113.19	159	Right	Occ	Early_Visual	V4_R	26.5	-81.6	-8.8	4.9295	0.0002
110.53	228	Left	Par	Inferior_Parietal	PGi_L	-48.4	-58.4	27	4.6878	0.0002
107.12	161	Left	Occ	MT+_Complex_and_Neighboring_Visual_Areas	V3CD_L	-37.8	-82.8	13.5	5.634	0.0002
105.53	192	Left	Par	Inferior_Parietal	PF_L	-55.1	-35.6	28.7	5.1037	0.0002
95.02	201	Right	Par	Temporo-Parieto-Occipital_Junction	PSL_R	58.3	-43.3	22.7	4.4061	0.0002
89	195	Left	Par	Superior_Parietal	VIP_L	-19.5	-64.9	56.8	4.2536	0.0002
88.28	175	Right	Fr	Inferior_Frontal	IFJp_R	33.6	5.6	31.4	3.9805	0.0002
86.02	135	Left	Temp	Lateral_Temporal	PHT_L	-54.2	-62.8	6.1	4.5631	0.0002
84.01	158	Left	Fr	Premotor	55b_L	-46.7	0.3	46.7	4.5385	0.0004
83.58	168	Left	Par	Inferior_Parietal	PGi_L	-43.4	-67.9	19.6	5.5873	0.0004
81.09	113	Right	Occ	MT+_Complex_and_Neighboring_Visual_Areas	MT_R	42.5	-73	0.7	3.9207	0.0006
79.95	128	Right	Fr	Premotor	55b_R	42.9	2.3	43	4.7482	0.0006
75.19	89	Left	Occ	Primary_Visual	V1_L	-15	-100.9	-4.8	4.2337	0.001
72.63	126	Right	Fr	Dorsolateral_Prefrontal	8C_R	37	19.2	26.7	4.2202	0.0006
68.81	117	Left	Fr	Dorsolateral_Prefrontal	p9-46v_L	-44.2	30.6	25.8	4.0961	0.0014

67.41	165	Left	Par	Inferior_Parietal	PFm_L	-48.4	-49	43.7	4.6699	0.0016
61.37	150	Left	Fr	Posterior_Opercular	43_L	-57.8	-0.7	14.5	4.6749	0.0024
55.58	83	Left	Occ	MT+_Complex_and_Neighboring_Visual_Areas	MT_L	-44.6	-72.1	6.3	4.5472	0.00499
54.71	112	Left	Fr	Premotor	6r_L	-46.6	2.4	26.3	3.6783	0.00559
54.6	71	Left	Temp	Ventral_Stream_Visual	FFC_L	-41.7	-64.7	-18.7	4.9128	0.00579
53.38	92	Left	Fr	Paracentral_Lobular_and_Mid_Cingulate	6ma_L	-19.8	-2.1	64.5	4.0547	0.00719
51.71	71	Right	Occ	Early_Visual	V4_R	25.9	-71.9	-5.4	5.8811	0.00898
50.52	115	Right	Par	Superior_Parietal	7Pm_R	8.8	-64.1	48.9	4.279	0.01017
49.89	99	Right	Par	Superior_Parietal	7PL_R	13.3	-71.6	52.2	4.3245	0.01057
48.04	123	Left	Par	Superior_Parietal	LIPd_L	-29.5	-47.9	41	3.9552	0.01435
47.54	111	Left	Par	Inferior_Parietal	PGs_L	-40.1	-63.9	33	4.5851	0.01534
46.66	81	Left	Par	Inferior_Parietal	IP1_L	-33	-66.3	42	5.061	0.01713
46.55	87	Right	Fr	Dorsolateral_Prefrontal	SFL_R	6.6	2.9	65.2	3.8731	0.01772
45.58	70	Left	Occ	MT+_Complex_and_Neighboring_Visual_Areas	V4t_L	-41.5	-77.1	-3.6	4.1141	0.01851
44.8	53	Right	Occ	Dorsal_Stream_Visual	V7_R	25	-75	27.5	4.0285	0.02247
43.37	76	Right	Fr	Anterior_Cingulate_and_Medial_Prefrontal	8BM_R	7	37.7	45.8	4.0643	0.02702
42.5	60	Right	Fr	Inferior_Frontal	44_R	47.2	20	9.4	3.817	0.03056
41.13	62	Right	Par	Inferior_Parietal	PGp_R	37.4	-79.1	13.6	4.0701	0.03489
40.42	47	Right	Occ	Dorsal_Stream_Visual	V7_R	23.7	-83.3	31.9	4.2199	0.03901
39.7	119	Right	Par	Posterior_Cingulate	PCV_R	7.5	-49.7	55.7	5.3128	0.04234
39.52	58	Left	Par	Inferior_Parietal	PGs_L	-35.1	-75.9	39.4	3.7602	0.04273
39.51	104	Left	Par	Inferior_Parietal	PF_L	-59	-31.3	36.7	3.9426	0.04273
38.84	95	Left	Par	Temporo-Parieto-Occipital_Junction	STV_L	-63.5	-41.9	11.1	4.8528	0.04645

Chapter V: General discussion

Humans constantly learn from their environment based on the observations they make. In a stochastic and dynamic environment, learning a hidden quantity (such as the probability of an event of interest occurring, or the expected size of a future reward) from observations is a difficult problem because the stochasticity makes the quantity impossible to determine with certainty and the dynamics imply that the quantity could have changed from any observation to the next. Classical learning theories rely on basic learning rules which are unable to solve this problem effectively and which incompletely explain human behaviour. In recent decades, researchers have begun to reframe learning theories within a normative probabilistic framework. In this framework, the learning problem is formalised as a sequential probabilistic inference problem that has an optimal solution derived from probability theory.

In my thesis, I analysed the predictions made by normative theory in different common learning contexts and identified properties and components of the normative solution that make the learning process adaptive and that humans should possess in order to learn effectively. In light of this analysis, I reviewed existing empirical and theoretical findings and identified several issues.

(1) Probability learning remained quite mysterious, particularly because the dynamics of the learning rate had never been examined in humans in this context, and because the normative determinants of these dynamics had not been well characterised.

(2) The feasibility in the brain of a solution possessing the identified normative adaptive learning properties, and the specific mechanisms that could enable these abilities in the brain, remained undetermined.

(3) The neural representation of probability, a fundamental component of learning, remained elusive.

I solved these issues through three studies which showed that:

(1) Probability learning, like magnitude learning, closely follows normative properties in humans. They dynamically and normatively adapt their learning rates, and differentially in the two contexts according to two normative determinants: surprise (or change-point probability), which plays a predominant role in magnitude learning, and uncertainty, which plays a predominant role in probability learning.

(2) Tiny recurrent neural networks, feasible in the brain, can perform quasi-optimal learning, possessing all the identified normative learning properties (including those demonstrated in study 1 in humans), provided that they have a gating mechanism.

(3) A nonlinear code represents probability in human brain activity within the dorsolateral prefrontal and intraparietal cortices.

Collectively, the results of these studies demonstrate the alignment of human learning with normative theory and highlight mechanisms and neural bases that could enable humans to use uncertainty and estimate probabilities as prescribed by the theory to achieve adaptive learning.

In this general discussion, I first discuss in section 1 points specific to each study, with a particular focus on limitations, and then in section 2, I discuss overarching points that highlight questions raised by my thesis and future directions.

1. Points specific to each study, focusing on limitations

Study 1

Magnitude learning and probability learning: How are they different? Implications of the differences we found

In study 1, we highlighted differences between magnitude learning and probability learning: the main factors driving the adjustments of the learning rate are not the

same. Surprise (or change-point probability) is the main driver of the adjustments in magnitude learning, while uncertainty is the main driver of the adjustments in probability learning. This leads to different profiles of learning rate dynamics: magnitude learning is characterised by sudden, immediate, and momentary adjustments of the learning rate following a highly surprising observation, while probability learning is characterised by gradual, more spread out adjustments of the learning rate covarying with uncertainty.

As we have shown, these results are expected according to the normative theory. Normative theory applies to both types of learning and allows us to establish a common model of learning. From this perspective, there are similarities between the two types of learning: they can both be treated analytically using the same formalism, and the observed behaviour in both cases adheres to the normative principles derived from this analysis. One could choose to focus mainly on these common aspects rather than on the differences. However I argue that neglecting these differences would be a potential mistake for understanding the learning process at play in the brain. To illustrate this, I will use an analogy between learning and visual recognition.

Two common types of visual recognition are face recognition and place recognition. The most relevant features for recognizing faces and places are not the same: for faces, it is better to pay attention to features such as eyes, nose, and corners of the mouth, while for places, it is better to pay attention to the global context (e.g., whether it is an indoor or outdoor place) and features such as the arrangement of buildings and roads, features which are larger-scale and have a different geometry than those for faces. It is now established that face recognition and place recognition are associated with two distinct brain structures: the fusiform face area (FFA), more specialised for faces, and the parahippocampal place area (PPA), more specialised for places (Epstein & Kanwisher, 1998; Kanwisher et al., 1997).

In both types of visual recognition, the problem to be solved can be described as that of associating a retinal image with a concept (a particular person, a particular place). This problem can be mathematically formalised, and has an optimal, normative, Bayesian solution. It is likely that one could find numerous behavioural matches

between the Bayes-optimal solution and human vision, given how well human vision can perform. However if one were to retain from those matches that visual recognition is implemented by a common, very general system, they would not necessarily be the most enlightened for understanding how the brain solves the problem. As illustrated by the research that ultimately led to the discovery of the FFA and the PPA, it is useful to focus on the different problem instances commonly encountered. By studying the different types of visual recognition, it is likely that one would find that normatively, a differential importance should be given to certain visual features for recognising faces versus places. Such a finding would indicate that there may be more specialised systems that the brain could employ to solve the problem effectively, and indeed it turns out that such systems are likely implemented by the FFA and the PPA.

Like visual recognition, it is possible that the brain relies on subsystems that are differentially solicited when learning magnitudes or probabilities. This is supported by the differences in the solicitation of surprise and uncertainty shown in study 1, and by the fact that surprise and uncertainty have distinct computational characteristics and partly distinct neural correlates (see section 4 of the general introduction). One way to determine this more clearly would be to conduct an experiment in which subjects would perform a magnitude learning task and a probability learning task as in study 1, but in a MRI scanner with simultaneous measurement of whole-brain functional activity.

Bias-variance decomposition at the subject level

In study 1, the bias-variance decomposition of subject deviations was performed at the group level. This informs us about the proportion of deviations attributable to biases that are reproducible across individuals. Such biases have been observed in other contexts than that of our study, as detailed in section 2 of this general discussion: biases in probability judgments (representativeness heuristic, sub- and super-additivity, partition dependence), biases in learning with decision-making (confirmation bias, positivity bias), as well as over- and under-weighting of low and high probabilities respectively as described in prospect theory (Kahneman & Tversky, 1979). In our study, the results of the group-level decomposition show that this type

of systematic biases constitute a relatively small part of the deviations compared to the variance of the subjects' estimates for the same sequence.

A remaining question is what the bias-variance decomposition is at the subject level. This question was not the focus of study 1 but is nevertheless interesting. It is possible that a subject has an estimation bias that is not visible at the group level because other subjects do not have the same bias. It is also possible that different subjects exhibit different biases in opposite directions (that is, at a given time step of a sequence, the biases of some subjects would lead to an overestimation and those of other subjects would lead to an underestimation). To perform the bias-variance decomposition at the subject level, one needs to estimate the variance of the estimates for the same sequence and the same subject. This could be done by presenting the same sequence at least twice to the same subject and collecting their estimates each time, as in (Drugowitsch et al., 2016).

Study 2

Interpretability of the network computations and the gating mechanism

I have proposed an understanding of the network computations through a detailed analysis of the networks' adaptive capabilities, the analysis of the mechanisms that allow the networks to achieve these capabilities (gating, lateral connections, tuning of the weights to the environment), a description of the internal activity dynamics in state space (Fig. 4b), the analysis of the variables represented in these internal activities (the hidden variables of the environment and associated uncertainties, disentangled from each other, Fig. 4a and 6d), and of their causal role (Fig. 5).

However, compared to a classical model whose algorithm has been designed by hand by researchers and is thus automatically interpretable, the algorithms realised by the networks are not as highly understood and one could hope to get even closer to such a degree of interpretability. The limitation of network size, which I was able to reduce to 2 or 3 recurrent units without substantially compromising performance (Fig. 8a), is a first step in this direction. A recent paper proposes, similarly, to use tiny RNNs, combined with state space analyses, as a way to increase interpretability (Ji-An et al., 2023). By performing a further model reduction, one could hope to arrive

at a simple and intelligible effective circuit like the ones offered by (Dubreuil et al., 2022) for solving simpler tasks. Another possible approach to improve network interpretability would be to introduce in the cost function used for training a term favouring interpretability, for example that would promote sparsity and penalise the amount of information contained in the network activities thus introducing an "information bottleneck" (Miller et al., 2023). These approaches overall are attempting to constrain the network in such a way that meaningful, interpretable internal variables are more likely to be extracted from the network. Other points of view exist on what constitutes a satisfactory explanation or understanding of computations that may not require interpretable internal variables to be extracted (Barack & Krakauer, 2021).

A more specific target for understanding is the gating mechanism. A more detailed understanding would help make the finding of its importance in adaptive learning even more interesting and useful for future research. The gating in the RNNs can be decomposed into several parts: there are two types of gates, the first type is applied to the connections upstream, and has a potentially different gating value for each connection, the other type is applied downstream after performing the summation and applying the non-linearity. Which parts are necessary? Are both types of gates necessary? Is it necessary to apply gating to all connections? Can a single gating value be used for all recurrent connections (versus connections from the input unit)? Simplifying gating to the maximum, following the same approach I used to simplify the architecture, would help creating closer links with biology in addition to helping understanding.

Ideally, one would also like to better understand why gating is necessary. I propose a first explanation: it adds the possibility of combining multiplicative operations with the additive operations of a simple recurrent network. This seems particularly useful for solving the adaptive learning problems insofar as the optimal solution given by Bayesian filtering combines additive operations (integrals) and multiplicative operations (see section 3.2 of the general introduction) (these are in general very frequent in probabilistic computations due to the use of the sum rule and the product rule of probability calculus). Necessity in a strict, mathematical sense remains to be demonstrated. There is considerable research currently being done in mathematics

to better understand which network architectures can efficiently solve which types of problems (by finding properties characterising approximations bounds and convergence as a function of network size) and which network architectures can or cannot be realised by a network of another architecture with limited size (examples of such properties have been found for deep versus shallow feedforward networks) (Barron, 1994; Eldan & Shamir, 2016; Lu et al., 2017).

Relating the networks with empirical data (behavioural or neural)

Study 2 is a theoretical study. The objective was to develop neural network-based learning models and to study them using simulations.

Another project would be to compare these networks with empirical data. Regarding behavioural data, a first dataset suitable to make such a comparison is now available: the one from study 1. It provides trial-by-trial behavioural data and thus allows for a fine comparison between network and human behaviour. The volume of data also allows for fitting the parameters of a network to the subjects' behaviour at the group level, and possibly at the subject level, using a small network (2 or 3 units) and few parameters.

Regarding neural data, the fMRI data used in study 3 could be used to test whether the activities of the recurrent units of a network can explain unique portions of neural activity in some regions using a voxel-wise encoding model (as in Study 3). This would involve inputting the sequence seen by the subject into the network, extracting from the network the sequence of internal activities (activations of the hidden units, and/or the gates if the network has some), and using these internal activities as regressors in the general linear model of the encoding model to predict the fMRI time series at the voxel level. Using artificial neural networks for the modelling of neural data has produced significant advances in predictive ability in the domains of visual perception, auditory perception, and language processing (Caucheteux & King, 2022; Eickenberg et al., 2017; Güçlü & van Gerven, 2015; Kell et al., 2018; Yamins et al., 2014). If the results obtained with RNNs were conclusive, it would be interesting to identify and quantify the importance of more specific factors for predicting neural data: the role of training the network for the task and of the amount of training, the role of the network architecture (with or without gating), the choice of internal

activities used to predict activity (hidden units or gates), the role of the input sequence (the one presented to the subject or another with certain statistics altered or preserved)... in order to gain insights into the information contained in the explained neural signal and the type of processing performed by the region producing that signal.

Such a comparison to empirical data would be all the more interesting that a maximum level of interpretability has been reached (see previous point in this discussion) and that the networks have been compared to other models and their differences characterised (see point about models in section 2 of this discussion).

Study 3

Differentiating neural signals encoding probability from those encoding outcome uncertainty

In the neural coding of probability we found, we observed a fairly high degree of symmetry with respect to 50%, meaning that probabilities of e.g. 30% and 70% were neurally encoded in a similar way. This degree of symmetry is mainly observed at the population level, notably through a representational similarity analysis (RSA): as quantified by the correlation across voxels of the neural responses, the similarity of the neural representations of p and $1-p$ is higher than one would expect from an encoding that represents probability with a similarity relationship that respects the order relation of the [0, 1] interval. If the encoding were purely symmetric with respect to 50%, it would be impossible to distinguish whether the encoded information is the probability (p) or the outcome uncertainty, as quantified, for example, by the Shannon entropy ($-p\log(p) - (1-p)\log(1-p)$) or the variance ($p(1-p)$) of the outcome implied by the probability. Although entropy was added as a covariate of the encoding model linear regression in our analyses, this method only controls for linear encodings of entropy, and it is possible that outcome uncertainty is encoded according to a nonlinear function of entropy. Due to the limited signal-to-noise ratio and the limited number of sessions available in the data for study 3, it is difficult to clearly differentiate a nonlinear encoding of probability from a nonlinear encoding of outcome uncertainty.

One possibility in the future to differentiate probability and outcome uncertainty would be to conduct a new experiment with three categories of outcomes instead of two. There would then be two probabilities to estimate (rather than just one), while outcome uncertainty would remain a single variable. One could then create conditions with different combinations of probabilities but equal outcome uncertainty, for example, the conditions 60/30/10%, 10/60/30%, and 30%/10%/60% (and in general, any permutation of the probabilities of the three categories). One would expect the neural signals to differentiate these conditions if they encode the probabilities of at least two categories in a stable manner (even if each probability was encoded with a symmetry with respect to 50% or 33%). (Note that the two probabilities may or may not be mixed in neural activity at the voxel level, this could be tested using conditions where all but one of the probabilities have changed.)

Apart from outcome uncertainty, another hypothesis that can explain the results we observed, is that the subject encoded the probability not as the probability of an arbitrarily chosen category (*p(A)*), but as the probability of the dominant category (*p(A)* when *p(A)* > *p(B)*, *p(B)* otherwise) combined with the identity of the currently dominant category (*A* or *B*). This is plausible because the stimuli used in the experiment for A and B were arbitrary (i.e., they did not have a very strong identity that would cause the subject to spontaneously treat *p(A)=70%* very differently from *p(B)=70%*), and because the scale displayed for the estimation response was visually represented as in that "dominant-probability" way: [A-100%-50/50%-100%-B].

The 3-category experiment proposed above would also allow us to distinguish between the hypothesis of encoding the two dominant probabilities versus the hypothesis of encoding outcome uncertainty and the hypothesis of encoding two category probabilities. To differentiate from outcome uncertainty, one could use, for example, the conditions 60/20/20% vs. 50/40/10%, which have almost equal entropies but different dominant probabilities. To differentiate from category probabilities, one could use conditions where the probabilities for two categories have been flipped, for example, 60/30/10% vs. 30/60/10%.

Relationship with probabilistic computations in the brain in general

In the task we studied, subjects explicitly (and covertly) estimated probability. This is not necessarily the case in all situations that involve probabilities, and assuming that the neural code for probability described in our study is involved in all these situations would be an overgeneralization. To contextualise our findings, let's consider other situations where the brain is performing probabilistic computations and is likely to represent probabilities in a different, more implicit way.

In situations where probabilistic contingencies are fixed or very stable, probabilities can be encoded not in neural activity but other biological substrates that are more stable and less energy-consuming, such as synaptic weights, or other durable storage formats in the brain (Gallistel, 2017; Hebb, 1949; Langille & Gallistel, 2020). Study 2 of this thesis provides an example, in a model network, of probability encoded in synaptic weights and used to optimally solve a task: the probability of a change point occurring (see Figure supplement 1 of article 2). In the brain, numerous associative learning experiments (e.g., fear conditioning, eyelid conditioning, appetite conditioning) have shown that the formation of associations occurs through neural plasticity in subcortical structures such as the amygdala, cerebellum, hippocampus, and brainstem structures (Fanselow & Poulos, 2005; Martin-Soelch et al., 2007; Thompson, 1988). When these associations are probabilistic, their probabilities are encoded in these structures since the probabilities dictate the degree of plasticity induced during learning. After learning, probabilities are implicitly taken into account in the computations performed by the brain to generate behaviour (e.g. the intensity or frequency of a conditioned response is proportional to the learned probability of an unconditioned stimulus coming next).

Another situation where the brain needs to perform more complex probabilistic computations than the previous one is probabilistic categorization tasks (Knowlton et al., 1994, 1996; Yang & Shadlen, 2007). In these tasks, on each trial, a combination of cues appears and a binary outcome is generated probabilistically with a probability calculated by combining the individual probabilities associated with each cue. The subject learns to associate the cues with probabilities through choices and trial and error: following the presentation of the cue combination, the subject chooses one of the two possible outcomes and receives feedback ("correct"/"incorrect," which may be associated with a material reward) depending on whether the true generated

outcome was the one chosen by the subject. This task amounts to an evidence accumulation task where each cue provides an independent sample of evidence. Indeed, after transforming probabilities (*p*) into log odds ratios (*log[p / (1-p)]*), the log odds ratio of the final probability is calculated by summing the log odds ratios associated with each cue. After learning, the brain has associated each cue with a value (a log odds ratio, or evidence strength) that represents a probability, and has learned to combine them according to the rules of probability calculus. In the experiment by Yang & Shadlen where this task was performed by rhesus monkeys, probability was reflected in the activity of LIP neurons, which increased or decreased following each cue depending on whether it favoured one outcome or the other, ultimately leading to the monkey's choice by a saccade (Yang & Shadlen, 2007). In humans, learning in this type of tasks is generally considered a procedural, implicit learning, and relies on cortico-striatal mechanisms (that are thus likely responsible for encoding cue probabilities) (Gluck et al., 2002; Knowlton et al., 1996).

Another situation is tasks involving sensory uncertainty. In cue combination tasks, for example, the subject must estimate a physical magnitude (e.g., height, position, orientation) based on multiple cues, each of which provides more or less precise information about that magnitude (e.g., an image of the object whose height must be judged, corrupted by noise) (Ernst & Banks, 2002; Körding & Wolpert, 2004). The information provided by the cue about the magnitude can be modelled by a probability distribution, and to make an accurate estimate, the brain must combine the information provided by the cues in accordance with the multiplication of their distributions, which amounts to giving more weight in the estimation to the cues that provide more precise information. Unlike the previous examples, in this type of task, the cues do not convey a probability per se but sensory information, and sensory uncertainty is often decoded from sensory cortices (Geurts et al., 2022; van Bergen et al., 2015; Walker et al., 2020).

In sum, all the situations above illustrate that there is a great diversity of ways in which the brain can perform probabilistic computations. The neural code for probability shown in study 3, where probability is explicitly encoded in neural activity and can be reported by the subject, has the advantage of being accessible and usable for many neural systems and cognitive tasks.

2. Overarching points, focusing on questions raised by the thesis and future directions

Reconciling our findings on normative human behaviour with those of the literature on human biases

If humans are capable of learning close to normatively, including probabilities (see article 1), how is it that biases are observed in humans when they judge probabilities or learn in other contexts?

Probability judgments and learning with decision-making are two domains where there is a large body of literature reporting human biases. Below, I detail these two cases one after the other. As I will explain, the results found in the literature on human biases are compatible with our results because the situations where these biases are observed are different from ours. By making the situations in which human biases occur more specific and highlighting specific factors that could cause the biases, our results actually help better understand human biases and provide clues as to their possible origins which could be investigated by future studies.

Probabilistic judgments. In probabilistic judgments, the following biases have been observed (for possible explanations of these biases, see e.g. (Costello & Watts, 2014; Zhu et al., 2023)).

> *Representativeness heuristic.* When people are asked to judge the probability that an object A belongs to category B (for example, judging the probability that "Steve is a librarian" based on a description of Steve as a "shy and withdrawn" person), the judged probability is strongly influenced by the degree to which A is judged to be similar to a prototype (i.e., representative) of B, and it is not sensitive enough to the prior probability of belonging to B (in the previous example, the prior probability corresponds to the proportion of librarians in the population) (Kahneman & Tversky, 1972). This bias can result in a *conjunction fallacy*, meaning that the probability that "A belongs to (B and C)" (e.g., that "Linda is a bank teller and a feminist") is judged to be greater than the probability that "A belongs to B" ("Linda is a bank teller") when the

description provided for A (Linda) resembles the prototype of C (a feminist) (Tversky & Kahneman, 1983).

Subadditivity. The probability that an event from a set A occurs (e.g., "dying from a natural cause") can be judged to be lower than the sum of the judged probabilities of A_1, A_2, and A_3, where A_1, A_2, and A_3 form a partition of A ("dying from a heart disease," "dying from cancer", and "dying from another natural cause") (Tversky & Koehler, 1994). This occurs when the unpacked examples for A (A_1 and A_2) are typical examples (i.e. with high probability). The reverse bias occurs (*superadditivity*) when the examples are atypical (i.e. with low probability) (Sloman et al., 2004).

Partition dependence. Probability judgments are biased depending on how events are partitioned within a set, in such a way that each element of a partition is judged to have a more equal probability, regardless of the partitioning. For example, the probability that "the temperature on Sunday will be higher than on any other day next week" is judged to be closer to 1/2, compared to the probability that "the day with the highest temperature next week will be Sunday", which is judged to be closer to 1/7 (Fox & Rottenstreich, 2003).

These biases in probabilistic judgments all share a commonality: they have been observed in situations where probability is judged based on a **description** (often verbal), rather than based on direct experience of the different possible events (as in study 1). This gap in subjects' behaviour depending on whether the behaviour is informed by descriptions or by experience has been observed repeatedly in decision-making (that is, description-based decision-making differs from experience-based decision-making) and is known in this field as the *description-experience gap* (Bévalot & Meyniel, 2023; Garcia et al., 2023; Hertwig & Erev, 2009). According to our results, it seems that, similar to decision-making, there is a useful distinction to be made between *probability judgment based on a description* and *probability judgment based on experience*.

Combining experience and description would be a worthwhile direction for future studies. In real life, our probability judgments are commonly influenced by both our

direct experience of the environment and the information communicated by others (social environment, media, etc.). For example, if one reads that a flu epidemic has been declared in their region, their judged probability of catching the flu will be influenced by both this information and their past experience of flu occurrences. In the laboratory, such an experiment could be done, for example, by asking the subject to perform a probability learning task like the one I developed for study 1, and inserting during the task some explicit information provided by a description (such as "the probability of blue is currently higher than 50%" or "there has been a change point in the last 5 observations"). The question is how subjects update their estimate after receiving such information.

Learning with decision-making. In learning with decision-making, biases have been observed in humans such as a *positivity bias*, where outcomes that have a positive valence for the subject are weighted more than those with negative valence (Frank et al., 2007; Lefebvre et al., 2017; Ting et al., 2022), and a *confirmation bias*, where outcomes that confirm the subject's choice are weighted more than those that disconfirm it (Palminteri, 2023; Palminteri, Lefebvre, et al., 2017; Talluri et al., 2018). The confirmation and positivity biases can both be formalised as a differential learning rate depending on whether the observed outcome induced a positively- or negatively-valenced prediction error and whether the outcome concerned the chosen or non-chosen option: for the chosen option, the learning rate is higher when the observed and obtained outcome is “better” than expected than when it is worse than expected, for the non-chosen option (for which the observed outcome is the one that the subject could have but did not obtain), the bias goes in the opposite direction (Palminteri & Lebreton, 2022).

The decision-making component seems to have a critical role in these biases. Indeed, the positivity and confirmation biases disappear when the subject does not freely make their own choices but instead the choices are imposed on them (free-choice condition versus forced-choice condition) (Chambon et al., 2020). Moreover, a large part of the biases explained by a differential learning rate can also be explained by a gradual perseveration in choices (i.e. a bias towards repeating previous choices) (Palminteri, 2023). Finally, a recent study demonstrated that certain biases in learning specifically occur when outcomes are produced by the subject's

choices (controllable condition), but not in the absence of choices when the subject only monitors the outcomes (uncontrollable condition) (Rouault et al., 2022).

The common denominator in all the above situations where biases have been observed in learning is that *the subject **makes choices*** and that *observations are perceived as consequences of their choices*. In this type of situation, the choice induces a polarity because some observations confirm and others disconfirm the choice made (which may be associated with a positive or negative valence, either because rewards are at stake or simply because of an intrinsic desire to be validated in one's choices). This polarity seems to give rise to asymmetries in human learning behaviour. In the studies I conducted, on the other hand, the observations from which the subject is learning do not result from their choices. This is a critical difference that likely explains why little to no biases have been observed in the studies of my thesis (at least that were reproducible across individuals, cf. article 1), while biases were observed in the other studies mentioned above.

In the future, it could be interesting to conduct experiments like those in the above-mentioned studies where, in addition to choices, estimates are requested from the subject, in order to investigate if the biases observed through the subject's choices are also observed in their estimates (it is possible that the two are not always consistent).

Testing the models on empirical data and drawing theoretical conclusions about human learning

I have listed computational models of learning in section 3.4 of the general introduction. Following study 2 of my thesis, the different architectures of recurrent neural networks that I have shown capable of learning should be added to this list.

One possibility in the future would be to test these models on behavioural or neural data and determine which model or models best explain the data. This comparison of empirical data to models and between models will be more interesting if theoretical insights can be drawn from the fact that one model or another better explains the data. What can these models and their comparison teach us about human learning?

An important theoretical analysis work is needed to clarify these potential insights. Since the models are numerous and their similarities and differences are not always clear, it would be useful to establish a categorization of the models, for example, according to the extent to which they exhibit different abilities (adaptations of the learning rate at different levels, taking into account uncertainty, etc.), according to the type of generative process they assume or inference process they perform (dynamics assumed to follow a random walk or a change point process, and whether change points are inferred, or volatility is inferred...), according to the variables they represent internally (e.g. uncertainty, probabilities, learning rate, samples from a distribution...) and the format in which they represent them (directly by an internal variable of the model, or encoded in or decodable from the model's variables, in a linear or non-linear way).

Having established useful categories of models, one needs to ensure that different models or categories of models can be clearly distinguished using empirical data. For this purpose, it would be wise to establish diagnostic tests, for example by using carefully chosen sequences and finding key distinctive behavioural effects, as in (Foucault & Meyniel, 2021; Heilbron & Meyniel, 2019; Palminteri, Wyart, et al., 2017). This is critical because two models that provide a priori very different interpretations can exhibit similar behaviours: for example, one that infers change points in an environment assumed to have change point dynamics, and another that infers volatility in an environment assumed to have random walk dynamics, where changes are assumed to occur on every time step, can both exhibit dynamic adjustments of the learning rate (see passage on the Piray and Daw model in section 3.4 of the general introduction). Conducting simulations, model recovery, and parameter recovery analyses is also necessary (Wilson & Collins, 2019). Simulating neural data can be challenging but regression, canonical correlation, and RSA between models can be used, and additionally, neural data can be generated using an encoding model as done in study 3.

In light of these theoretical analyses, the data available from existing studies (including those of this thesis) could be used to test the models. The insights drawn from model testing may be completely new compared to the existing study, or if a model had already been used in the existing study, they may reinforce the existing

results, or may propose a reinterpretation or a refinement of the existing results. For example, certain effects that were interpreted in terms of estimation of volatility could be interpreted differently in terms of stochasticity (Piray & Daw, 2021), or in terms of the width of the prior about the possible hidden variable values (Glaze et al., 2018), or in terms of change-point inference. New data could also be collected, informed by the theoretical analyses, in order to better distinguish models and maximise the insights gained about the human learning process.

Testing the hypothesis that the locus coeruleus-norepinephrine system regulates learning via a gating mechanism

I have shown in study 2 that the gating mechanism in artificial neural networks plays a key role in regulating learning. In the brain, such a gating mechanism can be biologically implemented by the locus coeruleus-norepinephrine (LC-NE) system. Indeed, gating is a mechanism by which the effective weight (or in other words, the *gain*) of a connection from an input to an output is modulated by a dynamic factor (the activity of a *gate* in the artificial network). It has been established that the LC-NE system performs analogous gain modulations in the brain. At the microscopic level, NE modulates the effective synaptic weight between neurons (it increases the neuronal responses evoked by synaptic inputs and reduces the background neuronal firing, thereby increasing the signal-to-noise ratio of the synaptic transmission between neurons) (Aston-Jones & Cohen, 2005; Berridge & Waterhouse, 2003; Rogawski & Aghajanian, 1980; Waterhouse & Woodward, 1980). NE can also have a cascading gain modulation effect as it can target neurons that release other neurotransmitters that also have a gain modulation effect at the synaptic level (Ferguson & Cardin, 2020). At the macroscopic, brain-wide level, gain modulation by the LC-NE system is observed through the behavioural responsiveness to meaningful events (Aston-Jones & Cohen, 2005; Donner & Nieuwenhuis, 2013). The observed modulatory actions of the LC-NE system are indeed well explained by a gain modulation/gating model (Servan-Schreiber et al., 1990).

In addition to the neural gating actions of the LC-system, the functional role of the LC-system in regulating learning is corroborated by existing theories (Aston-Jones &

Cohen, 2005; Yu & Dayan, 2005) as well as numerous empirical findings (see section 4 of the general introduction).

One hypothesis that follows from these different lines of work is that learning in the brain is regulated by a gating mechanism operated by the LC-NE system: the levels of LC activity and norepinephrine would be the biological analog of the gate activity levels in artificial neural networks. This hypothesis could be tested through new neuroimaging experiments involving learning. The following empirical predictions could be tested.

> *Correlation with behavior.* LC activity and NE levels should correlate with the subject's learning rate.
>
> *Correlation with gating (and gating-related) variables of the models.* LC-NE activity should be correlated with the activities of the gates in a recurrent neural network trained to perform the same task, or fitted to the subject's behaviour. It should also correlate with normative uncertainty (in results from study 2 that were not reported in the article, I found that uncertainty was even more accurately decoded from gate activities than from hidden unit activities).
>
> *Modulation of neural activity in other brain systems.* LC-NE activity should modulate neural signals of surprise, which can be measured by fMRI activations by surprise, stimulus-evoked pupillary and EEG/MEG responses, and the EEG/MEG signal power in certain frequency bands (Bounmy et al., 2023; McGuire et al., 2014; Meyniel, 2020). It should also modulate neural activity indexing the amount of update (Meyniel & Dehaene, 2017; J. X. O'Reilly et al., 2013), and possibly modulate the functional connectivity between regions encoding surprise and those encoding updates (Kao et al., 2020).
>
> *Causality.* Manipulation of LC-NE activity should systematically change the subject's learning rate, and change neural activity indexing updating for the same amount of surprise.

In practice, the experiments testing these predictions could combine a learning task with direct measurements of LC activity, using for example fMRI (de Gee et al., 2017;

Mazancieux et al., 2022), and simultaneous measurements of pupil diameter and of cortical activity using fMRI or EEG/MEG. NE levels can be manipulated pharmacologically using a selective norepinephrine reuptake inhibitor such as atomoxetine for example (Bymaster et al., 2002). Alternatively, a sensory stimulation that is irrelevant to the task could be investigated as an indirect but lightweight way to manipulate NE levels (Nassar et al., 2012).

References

Adams, R. P., & MacKay, D. J. C. (2007). *Bayesian Online Changepoint Detection*. https://arxiv.org/abs/0710.3742

Aston-Jones, G., & Bloom, F. E. (1981). Nonrepinephrine-containing locus coeruleus neurons in behaving rats exhibit pronounced responses to non-noxious environmental stimuli. *Journal of Neuroscience*, *1*(8), 887–900.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), Article 1. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Bach, D. R., & Dolan, R. J. (2012). Knowing how much you don't know: A neural organization of uncertainty estimates. *Nature Reviews Neuroscience*, *13*(8), Article 8. https://doi.org/10.1038/nrn3289

Baddeley, R. J., Ingram, H. A., & Miall, R. C. (2003). System identification applied to a visuomotor task: Near-optimal human performance in a noisy changing task. *Journal of Neuroscience*, *23*(7), 3066–3075.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329*(5995), 1081–1085.

Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature Reviews Neuroscience*, *22*(6), 359–371.

Barron, A. R. (1994). Approximation and estimation bounds for artificial neural networks. *Machine Learning*, *14*(1), 115–133. https://doi.org/10.1007/BF00993164

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., Shadlen,

M. N., Latham, P. E., & Pouget, A. (2008). Probabilistic Population Codes for Bayesian Decision Making. *Neuron*, *60*(6), Article 6. https://doi.org/10.1016/j.neuron.2008.09.021

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221.

Bendixen, A., Roeber, U., & Schröger, E. (2007). Regularity Extraction and Application in Dynamic Auditory Stimulus Sequences. *Journal of Cognitive Neuroscience*, *19*(10), Article 10. https://doi.org/10.1162/jocn.2007.19.10.1664

Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, *42*(1), 33–84.

Bévalot, C., & Meyniel, F. (2023). A dissociation between the use of implicit and explicit priors in perceptual inference. *bioRxiv*, 2023.08.18.553834. https://doi.org/10.1101/2023.08.18.553834

Bloom, L. (2013). *One word at a time: The use of single word utterances before syntax* (Vol. 154). Walter de Gruyter. https://books.google.com/books?hl=en&lr=&id=4bTnBQAAQBAJ&oi=fnd&pg=PA5&dq=Bloom+one+word+at+a+time&ots=Rmgwtx3xfL&sig=Ez6tpum7H-1LbH3mqfwveZnXA5I

Bounmy, T., Eger, E., & Meyniel, F. (2023). A characterization of the neural representation of confidence during probabilistic learning. *NeuroImage*, 119849.

Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology*, *58*(1), 49–67.

Browning, M., Behrens, T. E., Jocham, G., O'reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, *58*(5), 313.

Bymaster, F. P., Katner, J. S., Nelson, D. L., Hemrick-Luecke, S. K., Threlkeld, P. G., Heiligenstein, J. H., Morin, S. M., Gehlert, D. R., & Perry, K. W. (2002). Atomoxetine increases extracellular levels of norepinephrine and dopamine in prefrontal cortex of rat: A potential mechanism for efficacy in attention deficit/hyperactivity disorder. *Neuropsychopharmacology*, *27*(5), 699–711.

Caucheteux, C., & King, J.-R. (2022). Brains and algorithms partially converge in natural language processing. *Communications Biology*, *5*(1), 134.

Chaisangmongkon, W., Swaminathan, S. K., Freedman, D. J., & Wang, X.-J. (2017). Computing by robust transience: How the fronto-parietal network performs sequential, category-based decisions. *Neuron*, *93*(6), 1504–1517.

Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., & Palminteri, S. (2020). Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*, *4*(10), 1067–1079.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), Article 7.

Chen, Z. (2003). Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics*, *182*(1), 1–69.

Cook, J. L., Swart, J. C., Froböse, M. I., Diaconescu, A. O., Geurts, D. E., Den Ouden, H. E., & Cools, R. (2019). Catecholaminergic modulation of

meta-learning. *Elife*, *8*, e51439.

Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, *121*(3), Article 3. https://doi.org/10.1037/a0037010

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*(7), 294–300.

Courville, A. C., Daw, N., & Touretzky, D. (2004). Similarity and discrimination in classical conditioning: A latent variable account. *Advances in Neural Information Processing Systems*, *17*. https://proceedings.neurips.cc/paper_files/paper/2004/hash/65fc9fb4897a89789352e211ca2d398f-Abstract.html

Daw, N. D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*, *16*(2), 199–204.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876–879.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The helmholtz machine. *Neural Computation*, *7*(5), 889–904.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*(11), 1218–1223.

Dayan, P., & Long, T. (1997). Statistical models of conditioning. *Advances in Neural Information Processing Systems*, *10*. https://proceedings.neurips.cc/paper_files/paper/1997/hash/fe70c36866add1572a8e2b96bfede7bf-Abstract.html

de Gee, J. W., Colizoli, O., Kloosterman, N. A., Knapen, T., Nieuwenhuis, S., &

Donner, T. H. (2017). Dynamic modulation of decision biases by brainstem arousal systems. *eLife*, *6*. https://doi.org/10.7554/eLife.23232

de Xivry, J.-J. O., Coppe, S., Blohm, G., & Lefevre, P. (2013). Kalman filtering naturally accounts for visually guided and predictive smooth pursuit dynamics. *Journal of Neuroscience*, *33*(44), 17301–17313.

Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge University Press. https://cir.nii.ac.jp/crid/1130282271813126016

Dickinson, A., & Mackintosh, N. J. (1978). Classical conditioning in animals. *Annual Review of Psychology*, *29*(1), 587–612.

Donner, T. H., & Nieuwenhuis, S. (2013). Brain-wide gain modulation: The rich get richer. *Nature Neuroscience*, *16*(8), 989–990.

Douglas, R. J., & Martin, K. A. C. (2007). Recurrent neuronal circuits in the neocortex. *Current Biology*, *17*(13), R496–R500. https://doi.org/10.1016/j.cub.2007.04.024

Doya, K. (2008). Modulators of decision making. *Nat Neurosci*, *11*(4), Article 4. https://doi.org/10.1038/nn2077

Drugowitsch, J., Wyart, V., Devauchelle, A.-D., & Koechlin, E. (2016). Computational precision of mental inference as critical source of human choice suboptimality. *Neuron*, *92*(6), 1398–1411.

Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F., & Ostojic, S. (2022). The role of population structure in computations through neural dynamics. *Nature Neuroscience*, *25*(6), 783–794.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380.

Eickenberg, M., Gramfort, A., Varoquaux, G., & Thirion, B. (2017). Seeing it all:

Convolutional network layers map the function of the human visual system. *NeuroImage*, *152*, 184–194.

Eldan, R., & Shamir, O. (2016). The power of depth for feedforward neural networks. *Conference on Learning Theory*, 907–940. https://proceedings.mlr.press/v49/eldan16

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), Article 6870. https://doi.org/10.1038/415429a

Fan, H., Gershman, S. J., & Phelps, E. A. (2023). Trait somatic anxiety is associated with reduced directed exploration and underestimation of uncertainty. *Nature Human Behaviour*, *7*(1), 102–113.

Fanselow, M. S., & Poulos, A. M. (2005). The Neuroscience of Mammalian Associative Learning. *Annual Review of Psychology*, *56*(1), 207–234. https://doi.org/10.1146/annurev.psych.56.091103.070213

Ferguson, K. A., & Cardin, J. A. (2020). Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience*, *21*(2), 80–92.

Filipowicz, A. L., Glaze, C. M., Kable, J. W., & Gold, J. I. (2020). Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *Elife*, *9*, e57872.

Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., & Wyart, V. (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nature Neuroscience*, 1–12. https://doi.org/10.1038/s41593-019-0518-9

Fiorillo, C. D. (2003). Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. *Science*, *299*(5614), Article 5614. https://doi.org/10.1126/science.1077349

Fischer, A. G., & Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron*, *79*(6), 1243–1255.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences*, *14*(3), 119–130.

Foucault, C., & Meyniel, F. (2021). Gated recurrence enables simple and accurate sequence prediction in stochastic, changing, and structured environments. *eLife*, *10*, e71801. https://doi.org/10.7554/eLife.71801

Foucault, C., & Meyniel, F. (2023). Two determinants of dynamic adaptive learning for magnitudes and probabilities. *bioRxiv*, 2023–08.

Fouragnan, E. F., Chau, B. K., Folloni, D., Kolling, N., Verhagen, L., Klein-Flügge, M., Tankelevitch, L., Papageorgiou, G. K., Aubry, J.-F., & Sallet, J. (2019). The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. *Nature Neuroscience*, *22*(5), 797–808.

Fox, C. R., & Rottenstreich, Y. (2003). Partition Priming in Judgment Under Uncertainty. *Psychological Science*, *14*(3), 195–200. https://doi.org/10.1111/1467-9280.02431

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(41), Article 41.

https://doi.org/10.1073/pnas.0706111104

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127.

Gallistel, C. R. (2017). The coding question. *Trends in Cognitive Sciences*, *21*(7), 498–508.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review*, *107*(2), 289.

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review*, *121*(1), 96.

Garcia, B., Lebreton, M., Bourgeois-Gironde, S., & Palminteri, S. (2023). Experiential values are underweighted in decisions involving symbolic options. *Nature Human Behaviour*, *7*(4), 611–626.

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *37*(7), 1297–1310.

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*(1), 197.

Gershman, S. J., Pesaran, B., & Daw, N. D. (2009). Human reinforcement learning subdivides structured action spaces by learning effector-specific values. *Journal of Neuroscience*, *29*(43), 13524–13531.

Geurts, L. S., Cooke, J. R., van Bergen, R. S., & Jehee, J. F. (2022). Subjective confidence reflects representation of Bayesian probability in cortex. *Nature Human Behaviour*, *6*(2), 294–305.

Glaze, C. M., Filipowicz, A. L., Kable, J. W., Balasubramanian, V., & Gold, J. I. (2018). A bias–variance trade-off governs individual differences in on-line

learning in an unpredictable environment. *Nature Human Behaviour*, *2*(3), 213–224.

Glaze, C. M., Kable, J. W., & Gold, J. I. (2015). Normative evidence accumulation in unpredictable environments. *eLife*, *4*. https://doi.org/10.7554/eLife.08825

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the “weather prediction” task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory*, *9*(6), 408–418.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning book. *MIT Press*, *521*(7553), 800.

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F Radar and Signal Processing*, *140*(2), 107. https://doi.org/10.1049/ip-f-2.1993.0015

Greeno, J. G. (1980). Psychology of learning, 1960–1980: One participant’s observations. *American Psychologist*, *35*(8), 713.

Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005–10014.

Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory.* (pp. xix, 335). Wiley.

Heilbron, M., & Meyniel, F. (2019). Confidence resets reveal hierarchical adaptive learning in humans. *PLoS Computational Biology*, *15*(4), e1006972.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Horvath, J., Czigler, I., Sussman, E., & Winkler, I. (2001). Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. *Cognitive Brain Research*, *12*(1), 131–144.

Hoyer, P. O., & Hyvärinen, A. (2002). Interpreting Neural Response Variability as Monte Carlo Sampling of the Posterior. *Advances in Neural Information Processing Systems*, 2002.

Hunt, L. T., & Hayden, B. Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. *Nature Reviews. Neuroscience*, *18*(3), Article 3. https://doi.org/10.1038/nrn.2017.7

Iigaya, K. (2016). Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *eLife*, *5*. https://doi.org/10.7554/eLife.18073

Jepma, M., Murphy, P. R., Nassar, M. R., Rangel-Gomez, M., Meeter, M., & Nieuwenhuis, S. (2016). Catecholaminergic Regulation of Learning Rate in a Dynamic Environment. *PLoS Computational Biology*, *12*(10), Article 10. https://doi.org/10.1371/journal.pcbi.1005171

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, *23*(7), 1587–1596.

Ji-An, L., Benna, M. K., & Mattar, M. G. (2023). Automatic Discovery of Cognitive Strategies with Tiny Recurrent Neural Networks. *bioRxiv*, 2023–04.

Joshi, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences*, *24*(6), 466–480.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and

Cingulate Cortex. *Neuron*, *89*(1), Article 1. https://doi.org/10.1016/j.neuron.2015.11.028

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), Article 3. https://doi.org/10.1016/0010-0285(72)90016-3

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), Article 2. https://doi.org/10.2307/1914185

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), Article 1. https://doi.org/10.1115/1.3662552

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Kao, C.-H., Khambhati, A. N., Bassett, D. S., Nassar, M. R., McGuire, J. T., Gold, J. I., & Kable, J. W. (2020). Functional brain network reconfiguration during learning in a dynamic environment. *Nature Communications*, *11*(1), 1682.

Kell, A. J., Yamins, D. L., Shook, E. N., Norman-Haignere, S. V., & McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, *98*(3), 630–644.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*(12), 712–719.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A Neostriatal Habit Learning System in Humans. *Science*, *273*(5280), 1399–1402. https://doi.org/10.1126/science.273.5280.1399

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*(2), 106–120.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), Article 6971.

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nature Communications*, *6*(1), 8537.

Langille, J. J., & Gallistel, C. R. (2020). Locating the engram: Should we look for plastic synapses or information-storing molecules? *Neurobiology of Learning and Memory*, *169*, 107164.

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, *18*(8), Article 8. https://doi.org/10.1038/nn.4064

Lee, S., Gold, J. I., & Kable, J. W. (2020). The human as delta-rule learner. *Decision*, *7*(1), 55.

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, *1*(4), 0067.

Lieder, F., Daunizeau, J., Garrido, M. I., Friston, K. J., & Stephan, K. E. (2013). Modelling Trial-by-Trial Changes in the Mismatch Negativity. *PLoS Computational Biology*, *9*(2), Article 2. https://doi.org/10.1371/journal.pcbi.1002911

Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. (2017). The expressive power of neural networks: A view from the width. *Advances in Neural Information Processing Systems*, *30*.

https://proceedings.neurips.cc/paper_files/paper/2017/hash/32cbf687880eb1674a07bf717761dd3a-Abstract.html

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438.

Ma, W. J., Kording, K. P., & Goldreich, D. (2023). *Bayesian Models of Perception and Action: An Introduction*. MIT press.

Mackintosh, N. J. (Nicholas J. (1983). *Conditioning and associative learning*. Clarendon Press and Oxford University Press. https://cir.nii.ac.jp/crid/1130000795654535808

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-Trial Fluctuations in the Event-Related Electroencephalogram Reflect Dynamic Changes in the Degree of Surprise. *The Journal of Neuroscience*, *28*(47), Article 47. https://doi.org/10.1523/JNEUROSCI.2925-08.2008

Martin-Soelch, C., Linthicum, J., & Ernst, M. (2007). Appetitive conditioning: Neural bases and implications for psychopathology. *Neuroscience & Biobehavioral Reviews*, *31*(3), 426–440.

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00825

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A bayesian

foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, 39. https://doi.org/10.3389/fnhum.2011.00039

Mazancieux, A., Mauconduit, F., Amadon, A., de Gee, J. W., Donner, T., & Meyniel, F. (2022). Brainstem fMRI signaling of surprise across different types of deviant stimuli. *bioRxiv*, 2022–07.

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, *84*(4), 870–881.

Meyniel, F. (2020). Brain dynamics for confidence-weighted learning. *PLOS Computational Biology*, *16*(6), Article 6. https://doi.org/10.1371/journal.pcbi.1007935

Meyniel, F., & Dehaene, S. (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proceedings of the National Academy of Sciences*, *114*(19), E3859–E3868.

Meyniel, F., Schlunegger, D., & Dehaene, S. (2015). The sense of confidence during probabilistic learning: A normative account. *PLoS Computational Biology*, *11*(6), e1004305.

Miller, K. J., Eckstein, M., Botvinick, M. M., & Kurth-Nelson, Z. (2023). Cognitive Model Discovery via Disentangled RNNs. *bioRxiv*, 2023–06.

Murphy, P. R., Wilming, N., Hernandez-Bocanegra, D. C., Prat-Ortega, G., & Donner, T. H. (2021). Adaptive circuit dynamics across human cortex during evidence accumulation in changing environments. *Nature Neuroscience*, *24*(7), 987–997.

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal

systems. *Nature Neuroscience*, *15*(7), 1040.

Nassar, M. R., & Troiani, V. (2021). The stability flexibility tradeoff and the dark side of detail. *Cognitive, Affective, & Behavioral Neuroscience*, *21*, 607–623.

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, *30*(37), 12366–12378.

O'Reilly, J. X., Schuffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, *110*(38), Article 38. https://doi.org/10.1073/pnas.1305373110

O'Reilly, R. C., & Frank, M. J. (2006). Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation*, *18*(2), 283–328. https://doi.org/10.1162/089976606775093909

Ossmy, O., Moran, R., Pfeffer, T., Tsetsos, K., Usher, M., & Donner, T. H. (2013). The Timescale of Perceptual Evidence Integration Can Be Adapted to the Environment. *Current Biology*, *23*(11), Article 11. https://doi.org/10.1016/j.cub.2013.04.039

Palminteri, S. (2023). Choice-confirmation bias and gradual perseveration in human reinforcement learning. *Behavioral Neuroscience*, *137*(1), 78.

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., & Pessiglione, M. (2012). Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. *Neuron*, *76*(5), Article 5. https://doi.org/10.1016/j.neuron.2012.10.017

Palminteri, S., & Lebreton, M. (2022). The computational roots of positivity and confirmation biases in reinforcement learning. *Trends in Cognitive Sciences*. https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(22)00089-4

Palminteri, S., Lefebvre, G., Kilford, E. J., & Blakemore, S.-J. (2017). Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLoS Computational Biology*, *13*(8), e1005684.

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, *21*(6), Article 6. https://doi.org/10.1016/j.tics.2017.03.011

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532.

Piray, P., & Daw, N. D. (2020). A simple model for learning in volatile environments. *PLoS Computational Biology*, *16*(7), e1007963.

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications*, *12*(1), 6587.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160.

Prat-Carrabin, A., Wilson, R. C., Cohen, J. D., & da Silveira, R. A. (2021). Human Inference in Changing Environments With Temporal Structure. *Psychological Review*, *128*(5), 879–912. https://doi.org/10.1037/rev0000276

Pulcu, E., & Browning, M. (2017). Affective bias as a rational response to the statistics of rewards and punishments. *eLife*, *6*. https://doi.org/10.7554/eLife.27879

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, *2*, 64–99.

Ritz, H., Nassar, M. R., Frank, M. J., & Shenhav, A. (2018). A Control Theoretic Model of Adaptive Learning in Dynamic Environments. *Journal of Cognitive Neuroscience*, *30*(10), Article 10. https://doi.org/10.1162/jocn_a_01289

Rogawski, M. A., & Aghajanian, G. K. (1980). Modulation of lateral geniculate neurone excitability by noradrenaline microiontophoresis or locus coeruleus stimulation. *Nature*, *287*(5784), 731–734.

Rouault, M., Weiss, A., Lee, J. K., Drugowitsch, J., Chambon, V., & Wyart, V. (2022). Controllability boosts neural and cognitive signatures of changes-of-mind in uncertain environments. *ELife*, *11*, e75038.

Sahani, M., & Dayan, P. (2003). Doubly Distributional Population Codes: Simultaneous Representation of Uncertainty and Multiplicity. *Neural Computation*, *15*(10), 2255–2279. https://doi.org/10.1162/089976603322362356

Särkkä, S., & Svensson, L. (2023). *Bayesian filtering and smoothing* (Vol. 17). Cambridge university press.

Saxe, A., Nelli, S., & Summerfield, C. (2021). If deep learning is the answer, what is the question? *Nature Reviews Neuroscience*, *22*(1), Article 1. https://doi.org/10.1038/s41583-020-00395-8

Schultz, W. (2022). Dopamine reward prediction error coding. *Dialogues in Clinical*

Neuroscience.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. *Science*, *249*(4971), 892–895. https://doi.org/10.1126/science.2392679

Sheahan, H., Luyckx, F., Nelli, S., Teupe, C., & Summerfield, C. (2021). Neural state space alignment for magnitude generalization in humans and recurrent networks. *Neuron*, *109*(7), 1214–1226.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis.* https://psycnet.apa.org/Record/1939-00056-000

Skinner, B. F. (1953). *Science and human behavior.* https://psycnet.apa.org/record/1954-05139-000

Sloman, S., Rottenstreich, Y., Wisniewski, E., Hadjichristidis, C., & Fox, C. R. (2004). Typical versus atypical unpacking and superadditive probability judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 573.

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience*, 1.

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, *7*(2), 351–367.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential.

Science, *193*(4258), 1142–1146.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153.

Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? *Neural Networks*, *18*(3), 225–230.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), Article 11. https://doi.org/10.1038/nrn3838

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

Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M., & Donner, T. H. (2018). Confirmation bias through selective overweighting of choice-consistent evidence. *Current Biology*, *28*(19), 3128–3135.

Thompson, R. F. (1988). The neural basis of basic associative learning of discrete behavioral responses. *Trends in Neurosciences*, *11*(4), 152–155.

Ting, C.-C., Palminteri, S., Lebreton, M., & Engelmann, J. B. (2022). The elusive effects of incidental anxiety on reinforcement-learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(5), 619.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), Article 4. https://doi.org/10.1037/h0061626

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Tudusciuc, O., & Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *Proceedings of the National Academy of Sciences*, *104*(36), 14513–14518.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *101*(4), 547.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *24*(46), Article 46. https://doi.org/10.1523/JNEUROSCI.1905-04.2004

Vaghi, M. M., Luyckx, F., Sule, A., Fineberg, N. A., Robbins, T. W., & De Martino, B. (2017). Compulsivity reveals a novel dissociation between action and confidence. *Neuron*, *96*(2), 348–354.

van Bergen, R. S., Ma, W. J., Pratte, M. S., & Jehee, J. F. M. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience*, *18*(12), Article 12. https://doi.org/10.1038/nn.4150

Walker, E. Y., Cotton, R. J., Ma, W. J., & Tolias, A. S. (2020). A neural basis of probabilistic computation in visual cortex. *Nature Neuroscience*, *23*(1), 122–129.

Walker, E. Y., Pohl, S., Denison, R. N., Barack, D. L., Lee, J., Block, N., Ma, W. J., & Meyniel, F. (2023). Studying the neural representations of uncertainty. *Nature Neuroscience*. https://doi.org/10.1038/s41593-023-01444-y

Walton, M. E., Behrens, T. E., Buckley, M. J., Rudebeck, P. H., & Rushworth, M. F. (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron*, *65*(6), 927–939.

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., Hassabis, D., & Botvinick, M. (2018). Prefrontal cortex as a

meta-reinforcement learning system. *Nature Neuroscience*, *21*(6), 860.

Waterhouse, B. D., & Woodward, D. J. (1980). Interaction of norepinephrine with cerebrocortical activity evoked by stimulation of somatosensory afferent pathways in the rat. *Experimental Neurology*, *67*(1), 11–34.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158.

Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits*. Stanford Univ Ca Stanford Electronics Labs.

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife*, *8*, e49547.

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2013). A mixture of delta-rules approximation to bayesian inference in change-point problems. *PLoS Computational Biology*, *9*(7), Article 7. https://doi.org/10.1371/journal.pcbi.1003150

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624.

Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148), Article 7148. https://doi.org/10.1038/nature05852

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), Article 4. https://doi.org/10.1016/j.neuron.2005.04.026

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic Interpretation of Population Codes. *Neural Computation*, *10*(2), Article 2. https://doi.org/10.1162/089976698300017818

Zhu, J.-Q., Sundh, J., Spicer, J., Chater, N., & Sanborn, A. N. (2023). The autocorrelated Bayesian sampler: A rational process for probability judgments, estimates, confidence intervals, choices, confidence judgments, and response times. *Psychological Review*. https://psycnet.apa.org/record/2023-78840-001