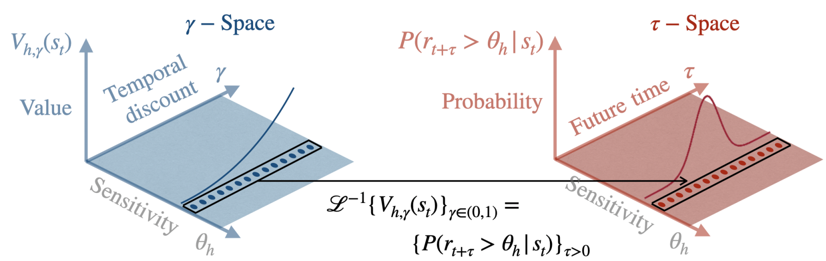
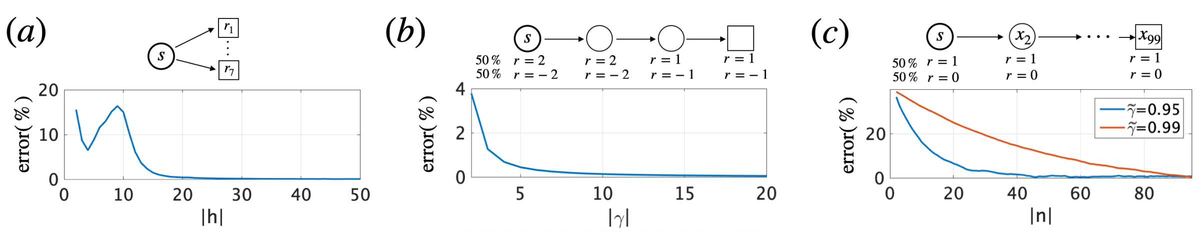
We are delighted to see that all reviewers recognized the important scientific contribution of our paper. We apologize for the dense writing and agree that we need to heavily edit the final version. As suggested by R2, we will add in particular a **background section**, which will include short description of traditional TD learning, the Laplace transform and the expectile code. The subsections on traditional TD learning and the Laplace transform should clarify Eq 2 and 3, respectively (as requested by R4). The nonlocality of the update rule of the expectile code (which was unclear for R2) will be dealt with explicitly in this section. Briefly, the update rule is non-local because to compute the RPE of unit *i*, unit *i* needs to know the state of a random unit *j*. This should make it easier to understand Section 4 even if the reader hasn't read [10]. We will also expand Section 2 to make it more gradual, stating theorems and definitions explicitly.

Figure 1 As noted by the reviewers, we reversed the figure captions for subplots b and c, for which we apologize. To clarify Fig. 1, we will insert the following diagram, which shows how we go back and forth, via the Laplace transform and its inverse, between the $V_{h,\gamma}$'s (the γ-space; left plot) and the temporal evolution of the immediate reward distribution (τ-space; right plot). The Markov Process in Fig. 1 (b) will then serve as an example for this diagram, clarifying that in the γ-space we now measure $V_{h-1,\gamma} - V_{h,\gamma}$ (instead of $V_{h,\gamma}$) to recover the distribution $P(r_\tau = \theta_h)$ (instead of the cumulative distribution, $P(r_\tau > \theta_h)$), a question from R4. We hope that this clarifies the connection between $\{V_{h,\gamma}\}_{\gamma \in (0,1)}$ and $\{P(r_{t+\tau} > \theta_h)\}_{\tau > 0}$, a concern of R3.



Range of h, γ, n and quality of approximation R3 expressed the concern that our approach might require an inordinate number of neurons in the Ventral Tegmental Area, where reward prediction errors are believed to be encoded in the brain. To investigate this issue, we explored how the number of units along the h-,γ-,and n-dimension govern the quality of the learnt value distribution. In ***(a)*** we show the normalized error in the γ-space (i.e. the percentage difference between the estimated and true $V_{h,\gamma}(s)$ distribution) using $|h|$ values of θ_h's uniformly distributed between r_1 and r_7 in the variable reward-magnitude task. In general, to represent P rewards with negligible error, we need $|h| \sim 3P$. In ***(b)*** we show the normalized error as a function of the number of γ's (using $|h| = 20$), for the Markov Process shown on the top. About 10 values of γ are sufficient, considerably less than the 300 values that we used in the original simulations (this result also holds in the τ-space, recovered with the linear decoder). Finally, in ***(c)*** we study the effect of $|n|$ in recovering the $V_{h,\gamma,n=99}(s)$ distribution (with $|h| = |\gamma| = 20$), for two different $\tilde{\gamma}$'s, the temporal discount defined by the problem. If $\tilde{\gamma}$ is not too high (< 0.95), the rewards received very far away in the future carry almost no weight, so including them (via a larger $|n|$) does not improve the estimated value distribution.



This analysis suggests that about 20 values covering the range along each dimension, for a total of 8000 units, can represent a wide range of problems. However, $|n|$ constitutes a very hard constraint on the range of representable problems, specially if $\tilde{\gamma}$ is high. This constraint softens with the continuous representation of the n-dimension (Eq. 8). Some other forms of function approximation could directly reduce the N^3 dimensionality, although possibly compromising locality. We will further explore the limits of representable problems in the final version.

Related work We agree with R2: the $V_{h,\gamma}$ to which the Laplace code converges can be formalized as a Generalized Value Function with cumulant function $C_t = f_h(r_t)$ and constant termination function γ. This interpretation opens interesting routes to function approximation, which we will discuss in the final version.

Code release We intend to release the code on GitHub as soon as the submissions are no longer anonymous.

Others We thank the reviewers for all the detailed comments (in particular R4). We can't respond to all of them here due to space constraints but we will make sure to address them all in the final version.