Optimal learning under structural environmental uncertainty reveals inherent learning trade-offs

Santiago Herce Castañón (S.HerceCastanon@GMail.Com)
Department of Psychology and Educational Sciences, University of Geneva, Switzerland

Pedro Cardoso-Leite (Pedro.CardosoLeite@Uni.Lu)
Department of Behavioural and Cognitive Science, University of Luxembourg, Belval, Luxembourg

C. Shawn Green (CShawn.Green@Wisc.Edu)
Psychology Department, University of Wisconsin-Madison, Madison, Wisconsin, USA

Daphne Bavelier (Daphne.Bavelier@UniGe.Ch)
Department of Psychology and Educational Sciences, University of Geneva, Switzerland

Paul Schrater (Schrater@UMN.Edu)
Department of Computer Science, University of Minnesota, Twin Cities, Minneapolis, Minnesota, USA

Abstract

In some contexts, human learning greatly exceeds what the sparsity of the available data seems to allow, while in others, it can fall short, despite vast amounts of data. This apparent contradiction has led to separate explanations of humans being equipped either with background knowledge that enhances their learning or with suboptimal mechanisms that hinder it. Here, we reconcile these findings by recognising learners can be uncertain about two structural properties of environments: 1) is there only one generative model or are there multiple ones switching across time; 2) how stochastic are the generative models. We show that optimal learning under these conditions of uncertainty results in learning trade-offs: e.g., a prior for determinism fosters fast initial learning but renders learners susceptible to low asymptotic performance, when faced with high model-stochasticity. Our results reveal the existence of optimal-paths-to-not-learning and reconcile within a coherent framework, phenomena previously considered disparate.

Keywords: optimal learning; volatility; background knowledge; structural uncertainty; prior for determinism

Introduction

One of the most striking aspects of human learning is the speed with which learning proceeds in some situations. Indeed, learning rates often exceed what the available data would allow, even if the learner was making maximum use of each data point (Carey & Bartlett, 1978; Chomsky, 1980; Feldman, 1997; Jern & Kemp, 2013; Ward, 1994; Xu & Tenenbaum, 2007). These findings suggest that humans use background knowledge when approaching novel learning problems (Lake et al., 2017; Tenenbaum & Griffiths, 2001). This type of generalization can allow an individual in a new learning situation to either show an immediate high level of performance, to have a faster learning rate, or both (Harlow, 1949; Kattner et al., 2017; Spelke et al., 1992). Yet, at the same time, human learning performance can sometimes be substantially lower than what the data would allow, even in seemingly simple tasks (Baker et al., 2014; Findling & Wyart, 2021; Wang et al., 2017; Wyart & Koechlin, 2016). Such failures of learning are typically attributed to completely different mechanisms than those supporting surprisingly fast learning. For instance, studies using sequence learning paradigms have found a proportion of the subjects to be "weak" learners of statistical properties of sequences (Baker et al., 2014) or to fall short of the expected optimal learning performance (Wang et al., 2017).

While current views rely on separate explanations for these two aspects of human learning: where "fast learning" assumes people use prior knowledge and optimal inference computations and "poor learning" assumes ad hoc suboptimal mechanisms (e.g., memory leakage, inattention, demotivation), we posit that these seemingly unreconcilable findings are in fact to be expected from optimal learning agents. First, the agents must have the ability to learn quickly (i.e., to discover the generative models, the causal explanations of their observations). This requires that learners have, from the start, a set of candidate generative models of the observations. Second, we posit that agents must use their observations to update their prior expectations of two sources of structural uncertainty about the environment: a) uncertainty about the stochasticity of each candidate model and b) uncertainty about the volatility of the environment. In other words, learners need to infer the extent to which the observations *must* be explained by their putative generative model, and the probability with which the underlying generative model could *switch* onto a different one from one moment to the next.

We argue that uncertainty about the structural properties of the environment in human learning tasks is much more pervasive than is currently recognised. Structural uncertainty is bound to occur in laboratory settings whenever instructions are unclear or incomplete, in ecological settings (where instructions are non-existent and goals may be unclear), and in many real-life scenarios where the context of the

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

environment is ambiguous (Acuña & Schrater, 2010; Beck et al., 2012; Behrens et al., 2007; Courville et al., 2006). Think for example about turning the radio on when an unknown jazz piece is playing. Try to predict what notes will be played next. Your predictions may differ, for instance, depending on whether there is a single player who is playing a recurring pattern of notes with some variations (i.e., random excursions from the main melody) or multiple players taking turns playing distinct melody pieces. The way we interpret this novel musical piece, and how well we are able to predict the forthcoming notes, will largely hinge on the nature of our background knowledge (based perhaps on similar prior experiences). At the core of this learning problem lies an inherent ambiguity for attributing the causes of unexpected observations. Should an unexpected stimulus be attributed to a switch in the true generative model (i.e., a switch to a new musician)? Or should it be attributed to a single non-deterministic true generative model (i.e., one musician who introduces stochastic variations)?

Here, we follow a Bayesian solution to learning under this type of structural uncertainty about the environment and report a series of counterintuitive learning trade-offs. First, a trade-off exists between: the speed of initial learning (and of adaptation to changes), and the ceiling or final performance under any one assumed model (Brand, 1999; Jaynes, 1982). We show that a learner's prior expectation towards deterministic models is simultaneously conducive to the fastest identification of the underlying model and of changes of the underlying models. However, such a prior for determinism also leads learners to quickly assume a change in models when in fact the observations may simply be stochastic departures within a single true model. Similarly, prior expectations for a stable (as opposed to volatile) environment lead to higher levels of final performance for all the observations that arise from a single model. However, such a prior for stability also precludes detecting possible changes in the underlying models. A second learning trade-off is linked to the size of a learner's set of candidate models. Intuitively, a bigger set of assumed models is more likely to contain the true model for a series of observations, and thus to yield higher performance levels. However, a bigger set of assumed models can make it harder to identify the true model (from a subset that makes few disjoint predictions), thus reducing the speed of learning. Together, we show that these learning trade-offs predict the existence of optimal-paths-to-not-learning for agents that aim to learn fast. Specifically, a prior for determinism can lead a learner to quickly identify the true model for a series of observations, and quickly identify a true change in model, at the cost of hindering learning when a single true model with higher stochasticity generates the observations.

Results

We analyse the learning consequences for agents that: i) have a set of candidate models for their observations (which allow them to learn at rates that exceed what the observations alone would allow), ii) have uncertainty about how stochastic each of the models may be and iii) have uncertainty as to the likelihood that a model can be overtaken in time by a different one. The last two points capture the agents' uncertainty about the structural properties of the environment. Although the general framework works for a broad set of learning environments, here we ground our work onto a specific case of learning in a sequence prediction task. Some of the assumptions made in the specific example are made for allowing completeness of information on the side of the learners, but we clarify what the minimal assumptions are that support generalizing our results.

Definitions, Notation and Assumptions

Briefly, we assume an environment that consists of the sequential presentation of one stimulus (at each time point) out of a fixed set (e.g., and unordered set of symbols). Agents have to predict at each time point what stimulus will come next, requiring in essence the agents to learn the predictive transition patterns. We describe the both the environment and the agents in more detail, and in turn, in the following subsections.

Environmental model The stimulus sequence is controlled by a specified set of generative models which probabilistically determine the next stimulus as a function of the last one. Each of the generative models is a first order Hidden Markov Transition model. Furthermore, the complete set of generative models that obey a set of rules form a generative family. In general, we assume that the size of the model space $N_m = |\mathcal{M}|$ is finite.

For grounding our results, we assume an environment where the generative family is defined by six possible generative models parameterised by α, a level o stochasticity. The six permutation matrices over a set of four symbols (four stimulus identities) span the full space stimulus-to-stimulus transitions and capture the most essential part of the generative models: the set of *dominant transitions*, i.e., the most likely next stimulus from any given stimulus. All dominant transitions for a model can be described by a dominant transition matrix W_P . A stochasticity parameter, α, describes the probability with which a dominant transition will occur under a given model. Within our chosen generative family, *self-transitions* do not occur (i.e., the stimulus is never the same on two consecutive time points). Finally, the two *non-dominant* transitions under a model for any given stimulus (i.e., all transitions that are not dominant or self-transitions) occur with equal probability (splitting the compliment of α over the two possibilities).

The environment is also characterised by a level of *volatility* which reflects the probability of switching between generative models across time. Importantly, the volatility is affected by the frequency of model switches and depends upon the number of models which can be visited in a given environment (i.e., six in our case).

Agent model An important conceptual distinction exists between an environmental model (i.e., a true model generating the stimulus) and an agent's assumed model of

how the stimuli are generated. However, we often use simply the word *model* when we mean either the *true model* of the environment or the agent's *assumed model*. Furthermore, and for completeness, we assume agents have the full set of generative models contained in the environment's generative family; yet, the generality of our findings hinges only on the agent having a set of two or more *distinct* models (i.e., assumed models that make *disjoint predictions*), and not necessarily on having the complete set of possible ones. We assume agents perfectly know the generative family at the level of the structural assumptions, that is, they know the possible dominant transition matrices, W_p, but they do not know the level of stochasticity in the models or the volatility in the environment. This assumption allows us to simplify the treatment of the agents' assumed models to be matched to the generative family but is not essential, and our results do not rely on the agents having perfect knowledge of the generative family (they rely solely on the agent having at least two competing candidate models as well as having uncertainty about the stochasticity of the models and having uncertainty about the volatility of the environment).

In general, we let $\theta_m = \{\alpha_1^m, \cdots \alpha_k^m\}$ be the set of k parameters in model m assumed by an agent. For the specific environmental model described in the previous section, we can approximate any of the models assumed by an agent as a transition matrix, T_{θ^m}, composed of a weighted sum of three parts: $T_{\theta^m} = \alpha_1 * (W_p) + \alpha_2 * D^{-1}(E - I - W_p) + \alpha_3 * (I)$, where W_p is any one of the six dominant transition matrices, I is a matrix that identifies all the transitions that cannot be observed within the environment (i.e., self-transitions), E is a matrix of all ones, such that $E - I$ identifies all transitions that can be observed with high likelihood under any of the models within the environment, and D^{-1} is a matrix that ensures the rows add up to 1, with the constraints that $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and that $\alpha_1 > \alpha_2 > \alpha_3 > 0$. Note that the actual transition probabilities of an environmental model need not coincide with the transition probabilities assumed by an agent. Finally, the parameters can be seen as mixing parameters that capture the assumed probability with which the stimulus, at each time point, will follow the dominant transition under the m^{th} model (α_1), will follow the dominant-transition under any of the other models (α_2), or will follow a transition for which no evidence has been given (α_3). When we use α without indexing we refer to α_1 for practicality.

We assume the process is fully observable (i.e., that the agent's observations correspond to the true stimulus identity), thus, we use *stimulus* and *observation* interchangeably. At any given time point t, the agent observes the stimulus $s(t)$ and is tasked with making a prediction $c(t)$ of what they think the next stimulus will be $s(t+1)$. Across all time points, the agent tries to maximize the total number of choices that match the forthcoming observation $c(t) == s(t+1)$. To achieve this, the agent needs to continuously infer: i) which model is in control of the stimulus, ii) the level of stochasticity of the models, and iii) the volatility of the environment. Despite the process being fully observable, a lapse rate (e.g., in the form of a noisy identity emission matrix) might be desirable. More concretely, if $\alpha_1 = 0.66$, $\alpha_2 = 0.32$ and $\alpha_3 = 0.02$, then for the following dominant transition matrix,

$$W_p^m = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

we can express an agent's model in the form of the following transition probability matrix:

$$T_{\theta^m} = \alpha_1 \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.02 & 0.16 & 0.66 & 0.16 \\ 0.16 & 0.02 & 0.16 & 0.66 \\ 0.16 & 0.66 & 0.02 & 0.16 \\ 0.66 & 0.16 & 0.16 & 0.02 \end{bmatrix}$$

Thus, the parameters of an agent's internal model simultaneously determine its predicted probability of observing each of the possible transitions of the stimuli, as well as each of the predictive choices from the agent (i.e., the predicted stimulus transition for the next trial).

Decision model Learning agents make *forecast choices* by forming the predictive distribution over next possible stimuli, given their model uncertainty. Given the focus on learning, these equations need to be written in *recursive form* to understand how choices vary in time as a function of history and priors.

We introduce two types of history—stimulus history and choice history. Specifically, $H_s(t) = \{s_0, s_1, \cdots, s_{t-1}\}$ is the history of stimuli up to, but excluding the t^{th} timestep, while $H_c(t)$ is the history of choices up to, but excluding, the t^{th} timestep. These equations are defined so that the histories at time t are consonant with the information the observer would have available for making their **forecast** of the stimulus $s(t)$, and allow us to handle the history and the latest stimulus (or choice) separately which simplifies our equations.

The agent maintains a dynamic belief about which model is currently controlling the stimulus, represented by a random variable z_t which indexes the model. The model belief is the probability $P(z_t = m)$. The agent also has an internal model for how the environmental model changes from one trial to the next, $V = P(z_t|z_{t-1})$ which represents their assumptions about the *volatility* of the environment (i.e., the probability that the model currently controlling the stimulus will continue to do so on the next trial, versus that other models will take over). To simplify, we parameterize this matrix with the continuity probability of ζ on the diagonal, and $(1-\zeta)/(N_m - 1)$ on the off-diagonal (i.e., all other models are equally likely to take over).

The forecast equation performs a weighted average of the models by their reliability to obtain its posterior probabilities:

$$\begin{aligned} Q(s_t|H_s(t)) &= P(s_t|s_{t-1}, \cdots, s_0) \\ &= \sum_m P(m|H_s(t))P(s_t|s_{t-1}, m) \end{aligned} \qquad \text{(eq. 1)}$$

The model posterior $K_t(m) = P(z_t = m|H_s(t))$ represents the *confidence* that the agent has on the model m being in control of the stimulus at time t. Note that the confidence is a normalized distribution over the set of models and will vary as a function of the history. The confidence in a model reflects its ability to explain the observed sequence of stimuli, given its parameter uncertainty. Our parametrization of

agents' assumed models allows us to expand the model posterior as a function of its parameters (see below).

Choices are modelled as a soft-max selection of the most likely prediction. We can write this in many equivalent forms. Using $log(Q)$ provides the most familiar form as:

$$P(c_t|s_{t-1}, H_s(t)) = \frac{exp(\gamma * log(Q(s_t = c_t|s_{t-1}, H_s(t))}{\sum_{s_t} exp(\gamma * log(Q(s_t|s_{t-1}, H_s(t))} \quad \text{(eq. 2)}$$

Particular choices of the agent are assumed to be sampled from this distribution (here, γ is a temperature parameter).

Confidence $K_t(m)$ is the most important indicator of the agent's understanding of the process at the level of *structural assumptions*. Each model has its own parameters, and the model's parameterization affects the confidence in important ways.

We can thus express confidence in a model in terms of the probability of the history (of stimuli and choices) under the model, as follows:

$$K_t(m) = P(z_t = m|H_s(t)) = \frac{P(H_s(t), z_t = m)}{\sum_{z_t} P(H_s(t), z_t = m)} \quad \text{(eq. 3)}$$

We can express the above equation, up to a normalization constant η, by expanding the stimulus history, $H_s(t)$ into two parts: the history up to the previous stimulus $H_s(t-1)$, and the latest stimulus $s(t)$:

$$K_t(m) = \eta P(s_t, H_s(t-1), z_t = m) \quad \text{(eq. 4)}$$

The above equation marginalizes over the possible values that the latent variable had in the previous trial z_{t-1}, by expanding it we obtain:

$$K_t(m) = \eta \sum_{z_{t-1}} P(s_t, H_s(t-1), z_t = m, z_{t-1}) \quad \text{(eq. 5)}$$

The expansion is useful because it allows us to express the confidence as the product of three components:

$$K_t(m) = \eta \sum_{z_{t-1}} P(s_t|s_{t-1}, z_{t-1}) P(z_t|z_{t-1}) P(z_{t-1}|H_s(t-1)) \quad \text{(eq. 6)}$$

The three components are: i) the confidence in the models at the previous timestep, i.e., the prior belief over models up to the previous timestep given the stimulus history at the previous timestep $P(z_{t-1}|H_{s(t-1)})$, ii) the model-to-model transition probability, i.e. the probability of a model at the current timestep given the model at the previous timestep $P(z_t|z_{t-1})$, and iii) the likelihood of the stimulus, i.e., the probability of the latest stimulus transition under the model at the previous timestep $P(s_t|s_{t-1}, z_{t-1})$.

We can rewrite the equation to express confidence as follows:

$$K_t(m) = \eta R_t(m) V K_{t-1}(m) \quad \text{(eq. 7)}$$

Where $K_{t-1}(m)$ is the confidence on the model at the previous timestep, V is the volatility matrix, and $R_t(m)$ is the responsibility that a model bears on having produced the latest stimulus.

Each model gets to marginalize over its internal parameters to make a forecast, so:

$$R_t = \int_{\theta_m} P(s_t|s_{t-1}, m, \theta_m) P(\theta_m|\{z_t, z_{t-1}, \cdots, z_1\}, H_s(t)) d\theta_t \quad \text{(eq. 8)}$$

Equation 7 shows that confidence at trial t, $K_t(m)$, is proportional (η) to the product of two components: the new evidence for the model, $R_t(m)$, and the previous confidence, $K_{t-1}(m)$.

We can now expand $K_t(m)$ using the model parameters. Here we make our first approximation, namely, that the parameter uncertainty's effect on the terms in $K_t(m)$ can be handled separately. This will be justified by conditioning on separate maximum likelihood estimates of the parameters for time steps t and $t-1$. This assumption bounds the true probability if the sequence of estimates converges. Given that the underlying process is stationary, convergence holds.

Our goal is to take the joint distribution on both z_t and on $s(t)$ and then unpack the model updates.

First, note that each transition model depends on its own parameters θ_m. Including those parameters, we write $P(s_t|s_{t-1}, z_{t-1})$ as $P(s_t|s_{t-1}, z_{t-1}, \theta_{z_{t-1}})$. To update the parameters, we form the posterior for each model given the history up to t. Assume we have an independent set of priors for each model $P(\theta_m)$.

The posterior probability of the m^{th} model after observing the latest stimulus is proportional to the likelihood of the stimulus transition under that model, conditioned on whether that model was active. Here we assume that z_{t-1} forms a "one-hot vector", which allows us to write the likelihood of the transition as:

$$P(s_t|s_{t-1}, z_{t-1}, \theta_{z_{t-1}}) = \prod_{m=1}^{N_m} P_m(s_t|s_{t-1}, \theta_m)^{z_{t-1}} \quad \text{(eq. 9)}$$

We can use Bayes' rule to express the probability of the model parameters as a function of the stimulus history (partitioning history into the last stimulus and previous stimuli as before). Let the set of parameters for all models be $\Theta = \{\theta_1, \cdots, \theta_{N_m}\}$. Then,

$$P(\Theta|H_s(t)) \propto \prod_{m=1}^{N_m} P_m(s_t|s_{t-1}, \theta_m)^{z_{t-1}} P(\theta_m|H_s(t-1)) P(z_{t-1}|H_s(t-1)) \quad \text{(eq. 10)}$$

Note that this equation shows that if the z probability is concentrated on one model at the previous timestep, then only that model is updated upon observing the new stimulus and all other models maintain their posteriors over parameter values. Model uncertainty thus "gates" parameter updates according to the model's responsibility for previous transitions. Despite the gating, the model is explicitly full memory—it does not include any forgetting. Forgetting can be modelled by adding transition dynamics on the model parameters. None of our key predictions require this complication.

The update equation can be parameterized in order to provide better intuitions for how the key concepts affect model learning. First, we can distinguish the first model as the true generator (of a series of stimuli generated by a single true generative model). In accordance with the true generative family of our environment, for dominant transitions $P_1(s_t|s_{t-1}, \theta_1) = \alpha$, self-transitions are zero, and all

others are $P_1(s_t|s_{t-1},\theta_1) = (1-\alpha)/(ns-2)$, where ns is the number of unique stimuli. Assuming a Dirichlet form for the prior, $P(\theta_1|H_s(t-1)) = \alpha^{n_1}(1-\alpha)^{m_1}$ for some positive integers n_1, m_1. Let,

$$M_{t-1} = \left(\prod_{m=1}^{N_m} P_m(s_t|s_{t-1},\theta_m)^{z_{t-1}} P(\theta_m|H_s(t-1))\right) \quad \text{(eq. 11)}$$

Then, the update for dominant transitions is:

$$\begin{aligned} P(\Theta|H_s(t)) &= M_{t-1}\left[\alpha^{z_{t-1}(1)}\alpha^{n_1}(1-\alpha)^{m_1}\right] P(z_{t-1}|H_s(t-1)) \\ P(\Theta|H_s(t)) &= M_{t-1}K_{t-1}(2:m)\left[\alpha^{n_1+z_{t-1}(1)}(1-\alpha)^{m_1}\right] K_{t-1}(1) \end{aligned} \quad \text{(eq. 12)}$$

In expectation, the count n_1 is updated by $P(z_{t-1}(1)|H_s(t-1))$ the expected value of the first component of z_{t-1}. *This means the effective rate of model learning is proportional to confidence, in other words learning is gated by confidence.*

The effects of priors on learning There are several ways to see how learning depends on the priors. Here we start by rewriting the mixture dynamics as a larger Markov model and then use Large Deviations Theory to get a convergence rate.

First we enlarge the state space $q_t = s_t \otimes z_t$ which is a $4N_m$ sized state space. The overall transition matrix is $T_q = T_{1:m} \otimes V$, which provides the joint $P(s_t, z_t|s_{t-1}, z_{t-1}) = T(s_t|s_{t-1}, z_{t-1}) * V(z_t|z_{t-1})$. This can also be viewed as a hidden Markov model, with observation $o_t = s_{t-1} \rightarrow s_t$, such that $T(s_t|s_{t-1}, z_{t-1})$ is the emission matrix and $V(z_t|z_{t-1})$ is the hidden transition matrix. In the hidden Markov model formalism, the emission is a doublet s_{t-1}, s_t and we track counts over these observed transitions.

Given this larger matrix, the empirical cumulative count induces a probability measure over the observations that converges to the true distribution. The convergence (learning) rate can be shown to be the differential relative entropy between the true distribution and the prior. This acts as the time constant of learning.

Simulations

We ran simulations to illustrate how learning depends upon the structural assumptions of the learner. An environment can be described by a pair of environmental parameters: i) the environment volatility (i.e., the probability with which the model switches onto a different one from one time point to the next), and ii) the model stochasticity (i.e., the probability that within a model an observation will occur that departs from the dominant transitions). Our simulated learning agents are also defined by two parameters (conceptually related to the environmental parameters): i) a prior for model determinism α (i.e., the agent's prior belief about the model stochasticity), and ii) the prior for volatility, ɣ (i.e., the agent's prior belief about the frequency switches of model). Note that the agent's belief may not always be consistent with the true structural properties of the environment in which it is trying to learn. We ran a series of simulations to show the effects of agents' prior beliefs on learning trajectories (**Fig. 1).** For each simulated learning agent (defined by its values of α and ɣ) and for each simulated environment, we ran one thousand independent and random instantiations of sequences and computed the corresponding learning trajectories (following equations above). We report the average performance across the different instantiations.

First, we show that a learner's prior for determinism is directly related to the speed with which they will converge onto (i.e., learn) the true generative model of a series of observations (**Fig. 1A)**. For these simulations, each simulated sequence started by choosing randomly (i.i.d) one of the possible environmental models as the true generative model for that sequence's complete series of deterministic observations. All instantiated learners started with a uniform prior over the possible assumed generative models, and knew that there was a single generative model for all observations (i.e., associated with ɣ=1) but we systematically varied the learners' prior for determinism (one hundred different and equidistant α values between the values of 1/3 and 1). Critically though, learning trajectories are not solely determined by the agents' prior for determinism and prior for volatility, but also by the environmental properties and the interactions between the them (**Fig. 1B)**. We focus on two prototypical environments (often used in laboratory settings) that lie along each of the axes of the space of environmental properties: i) a "prototypical single task learning paradigm" (big blue open circle), and ii) a "prototypical task switching paradigm" (big red open circle).

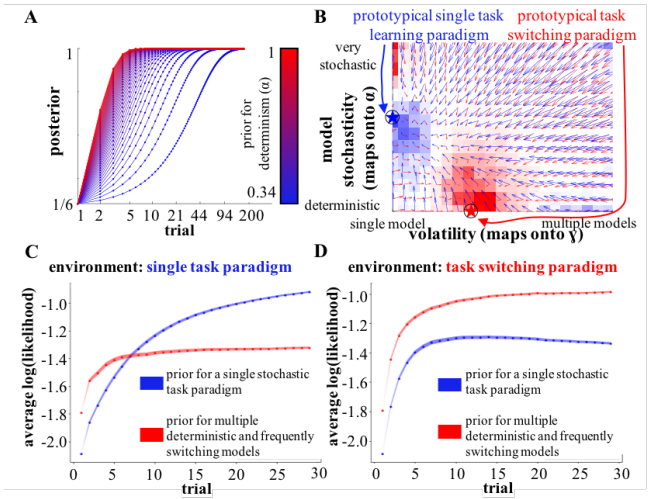

Figure 1: Trade-offs as a function of candidate models, beliefs in model stochasticity and environment volatility during optimal learning. (A) Convergence speed is dominated by the prior for determinism (α): a stronger prior for determinism (redder curves) results in sharper identification of the true underlying cause of a series of observations. (B) The structural properties of an environment can be described in terms of the "model stochasticity" and of the "environmental volatility" (number of models that can be switched onto and frequency of switching). Two prototypical environments frequently used in laboratory settings: i) a "single task learning paradigm" (blue star along y-axis) where there is a single true model to be learned and observations can stochastically differ from the model's

dominant transitions; and ii) a "task switching paradigm" (red star along y-axis) where multiple models deterministically generate the observations but frequent switches of the models occur. Learners' priors (arrow tails, projected in the environment for which they are best adapted) result in specific paths of learning (coloured arrows in the direction of updating when immersed in the environment). The colour of learners denotes the environment where they are immersed. The expected convergence solution depends on the region where the initial prior falls. Shaded coloured regions mark regions of near-zero expected update for agents in each environment. (C) In a "single task learning" environment, a trade-off exists between the priors that allow a fast identification of the underlying cause and susceptibility, in the long-run, to the stochasticity of observations. (D) The same priors that protect learners from stochasticity in observations ("a prior for a single stochastic task", blue curve), can prevent them from adapting to actual switches in the models that generate the observations. A learner's prior is parameterised as a Dirichlet distribution (for visualization only, we index learners by the max of their prior distribution).

Learners and their expected learning trajectories (and gradients) can be projected onto the same space for which their priors are best adapted. Learning depends both upon the environment in which they are immersed (difference between blue and red arrows) and upon the prior with which learners start. Ideally, learners immersed in an environment should converge to the solution maximally aligned with the environment, marked with the blue and red star for the "single task learning paradigm" and the "prototypical task switching paradigm", respectively. While learning trajectories are influenced by the actual stochasticity of the sampled stimulus (not shown), we can also compute the expected average gradients. The blue and red shaded areas represent expected gradients that are close to zero in either environment. Indeed, the ideal solutions for each environment are surrounded by shaded areas of their respective colours, revealing attractors. Attractors are a consequence of learning being gated by confidence and the preferential attachment of observations onto generative models (Equation 12), which can lead some learners with a "wrong" prior to convergence onto a solution that is wrong for the environment (i.e., red shaded area along the top part of the y-axis and blue shaded area towards the right part of the x-axis).

We then focus on the two prototypical environments and show the learning trajectories of optimal learners that have prior beliefs that correspond with either the properties of the true environment in which they are immersed or the other one (**Fig. C-D**). We simulated one thousand sequences for both a "prototypical single task learning" environment (environment parameters: α=0.75, γ=1) and for a "prototypical task switching" environment (environment parameters: α=1, γ=0.75). Assuming a single stochastic generative model for all observations (e.g., α=0.75 and γ=1; blue curves) conduces learners to performance which is resilient to unexpected observations and is desirable in environments that are truly controlled by a single generative model (**Fig. 1C**), but is maladaptive in environments where the generative models of observations are frequently switching (**Fig. 1D**). On the other hand, assuming multiple and frequently switching deterministic models (e.g., α=1 and γ=0.75; red curves) will allow for quick identification of the switches (**Fig. 1D**) but can be maladaptive in environments where a single stochastic model generates observations (**Fig. 1C**).

Discussion

Here we show the impact of an intrinsic ambiguity that optimal learners must navigate when interpreting the structural causes of stochastic observations. Together our results show that the prior beliefs of a learner regarding the properties of the environment, will result in learning that trades-off: i) quick learning and quick identification of switches, at the expense of vulnerability to within-model stochastic observations, against ii) resilience to stochastic observations which allow for high sustained performance in environments with a single stochastic cause for observations (at the expense of slower learning and identification of the dominant underlying cause in volatile environments).

We note that our treatment of volatility is not the same as a related concept previously introduced (Behrens et al., 2007; Piray & Daw, 2021; Wilson et al., 2010); in our case, high volatility is not only related to an increased probability of a switch in the cause but also to a higher number of causes that can be switched onto. We further show that when there is uncertainty about the structural properties of the environment, optimal learning can lead, through gated learning, to convergence solutions that may not be well-matched with the true properties of the environment. The converged solution for a learner will largely depend upon its background knowledge (i.e., its initial prior understanding of the environmental properties). We thus explain variability in learning outcomes in terms of variability in priors, instead of relying on ad-hoc suboptimal mechanisms like imperfect computations (Findling & Wyart, 2021) or probability matching (Acerbi et al., 2014; Shanks et al., 2002), without recognising the uncertainty about the structure of the environment (i.e., stochasticity vs volatility) that may have riddled the learners.

We endow our agents with background knowledge to allow them for the quick learning that characterizes humans in some contexts. Our approach is agnostic as to how the agents acquired background knowledge. Some aspects of background knowledge have been shown to be mostly learnt from previous experiences (Jusczyk, 2003; Madole & Cohen, 1995; Smith et al., 2002). Accordingly, in a recent study (Castañón et al., 2021), humans were shown to inductively infer aspects of the generative family with only a handful of trials. Yet, background knowledge can also be include an important innate component (Chomsky, 1980; Keil & Sessar, 1979; Spelke et al., 1992). While, our equations express agents that perfectly know the generative family from which

generative models are drawn, our results hold without this assumption; the critical assumptions are that learners: i) must have *a set of candidate models even if incomplete, and ii) must be uncertain about the stochasticity of models and about the volatility of the environment.* Our work takes an optimal modelling approach to human learning when allowing agents to include background knowledge that allow for the quick learning that characterizes human learning in some settings. In doing so, our work also provides a rational explanation for the failures of learning that characterizes humans in other settings. In particular, a background knowledge composed of a suitable set of candidate models and a prior bias for determinism can account for both aspects of learning without the need to invoke suboptimal mechanisms.

An exciting avenue of future research is to understand the mechanisms that drive humans to make early and fast inductive inferences such that their background knowledge can be suitably adapted to the environmental demands.

Acknowledgments

We want to thank Aaron Cochrane for useful comments on the manuscript, and Amanda Yung for help with programming the tasks. This work was supported and funded by: a Swiss National Science Foundation grant: 100014_15906/1 to DB; the Luxembourg National Research Fund: ATTRACT/2016/ID/11242114/DIGILEARN to PCL; the INTER Mobility/2017-2/ID/11765868/ULALA to PCL and PS. The Office of Naval Research grant N00014-17-1-2049 to CSG. The Office of Naval Research grant MURI GRANT N00014-07-1-0937 to DB.

References

Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLoS Computational Biology*, *10*(6), e1003661.

Acuña, D. E., & Schrater, P. (2010). Structure learning in human sequential decision-making. *PLoS Comput Biol*, *6*(12), e1001003.

Baker, R., Dexter, M., Hardwicke, T. E., Goldstone, A., & Kourtzi, Z. (2014). Learning to predict: Exposure to temporal sequences facilitates prediction of future events. *Vision Research*, *99*, 124–133.

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not noisy, just wrong: The role of suboptimal inference in behavioral variability. *Neuron*, *74*(1), 30–39.

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.

Brand, M. (1999). Pattern discovery via entropy minimization. *AISTATS*.

Carey, S., & Bartlett, E. (1978). *Acquiring a single new word.*

Castañón, S. H., Cardoso-Leite, P., Altarelli, I., Green, C. S., Schrater, P., & Bavelier, D. (2021). A mixture of generative models strategy helps humans generalize across tasks. *BioRxiv*.

Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, *3*(1), 1–15.

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.

Feldman, J. (1997). The structure of perceptual categories. *Journal of Mathematical Psychology*, *41*(2), 145–170.

Findling, C., & Wyart, V. (2021). Computation noise in human learning and decision-making: Origin, impact, function. *Current Opinion in Behavioral Sciences*, *38*, 124–132.

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, *56*(1), 51.

Jaynes, E. T. (1982). On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, *70*(9), 939–952.

Jern, A., & Kemp, C. (2013). A probabilistic account of exemplar and category generation. *Cognitive Psychology*, *66*(1), 85–125.

Jusczyk, P. W. (2003). Chunking language input to find patterns. *Early Category and Concept Development*, 17–49.

Kattner, F., Cochrane, A., Cox, C. R., Gorman, T. E., & Green, C. S. (2017). Perceptual learning generalization from sequential perceptual training as a change in learning rate. *Current Biology*, *27*(6), 840–846.

Keil, F. C., & Sessar, K. (1979). *Semantic and Conceptual Development: An Ontological Perspective*. Harvard University Press.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*.

Madole, K. L., & Cohen, L. B. (1995). The role of object parts in infants' attention to form-function correlations. *Developmental Psychology*, *31*(4), 637.

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications*, *12*(1), 1–16.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*(3), 233–250.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*(1), 13–19.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*(4), 605.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640.

Wang, R., Shen, Y., Tino, P., Welchman, A. E., & Kourtzi, Z. (2017). Learning predictive statistics from temporal sequences: Dynamics and strategies. *Journal of Vision*.

Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, *27*(1), 1–40.

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2010). Bayesian online learning of the hazard rate in change-point problems. *Neural Computation*, *22*(9), 2452–2476.

Wyart, V., & Koechlin, E. (2016). Choice variability and suboptimality in uncertain environments. *Current Opinion in Behavioral Sciences*, *11*, 109–115.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245.