Deep bidirectional interplay between sensory integration and latent rule discovery

Lucas Benjamin[1,2], Benjamin Morillon*[1], Valentin Wyart*[2]

[1] Aix Marseille Université, INSERM, INS, Institut de Neurosciences des Systèmes, Marseille, France
[2] Département d'Etudes Cognitives, École Normale Supérieure, Paris, France
* shared senior authorship

Abstract

The ability to decipher the structure of sensory input is crucial for building an accurate internal model of one's environment. Humans can not only extract and integrate the sensory properties of stimulus sequences but also discover hidden rules governing their generation. However, sensory integration and the discovery of latent generative rules in sequences have been studied separately, and whether and how these two cognitive abilities interact remains unclear. We jointly investigated sensory integration and rule discovery in a visual prediction task where a hidden rule governed predictable changes in the sensory properties of presented stimuli – a rule that only half of tested participants discovered. Participants' performance in sensory integration during the control blocks (without a hidden rule) accurately predicted their ability to discover the rule. Conversely, discovering the rule modified participants' sensory integration strategy. Using computational modeling of human behavior and simulations of recurrent neural networks, we show that minimizing leak during sensory integration is the sole parameter with a selective causal impact on subsequent rule discovery. Furthermore, participants who discovered the hidden rule appeared to apply it mentally to each presented stimulus, prior to integrating it. Together, our results reveal a deep bidirectional interplay between sensory integration and latent rule discover, to form a comprehensive learning system.

Introduction

Humans' sensory environment is uncertain, yet highly organized. Understanding the generative rules that govern its structure is crucial for reducing perceived uncertainty. Since sensory events typically unfold in time, researchers have extensively studied perception as a sequential process that extracts not only the sensory properties of sequences of stimuli, but also discovers latent rules that describe their temporal structure *(Bogacz, 2007; Clegg et al., 1998; Dehaene et al., 2015)*. However, and perhaps surprisingly, the cognitive mechanisms of sensory integration and rule discovery have historically been studied separately. Research on sensory integration focuses on how decision-relevant information is accumulated across streams of stimuli to reach a decision *(Ma et al., 2023; Ratcliff and McKoon, 2008; Summerfield and Egner, 2009; Wald and Wolfowitz, 1948)*. This sensory integration process is subject to a number of suboptimalities which, taken together, set an upper bound on the accuracy of human decisions *(Beck et al., 2012; Drugowitsch et al., 2016; Findling and Wyart, 2021; Rahnev and Denison, 2018)*. However, it remains largely unknown whether these suboptimalities only affect sensory integration itself, or whether they act as an information bottleneck on higher-order cognitive mechanisms.

To investigate this possible 'snowball effect' of integration suboptimalities on higher-order cognitive abilities, we designed a visual prediction task involving both sensory integration and the discovery of a hidden rule. Rule discovery in sequences is another well-studied human ability, corresponding to the identification of predictable temporal rule structures in sequential patterns (such as ABABAB… or AABBAABB… for sequences of binary stimuli). The ability to discover such rules is determined by their algorithmic complexity *(Dehaene et al., 2022; Quilty-Dunn et al., 2022)*. However, in most rule discovery tasks, sequences of stimuli are typically stripped of moment-to-moment randomness, eliminating the need to integrate sensory information across stimuli, and preventing any possible interplay between sensory integration and rule discovery. Here we designed a visual prediction task based on sequences of oriented stimuli sampled from probability distributions whose statistics change at a specific point in time in the sequence according to a simple latent rule. Unlike previous work, this task requires both efficient sensory integration and the discovery of the hidden rule to achieve accurate predictions.

A key distinction between sensory integration and higher-order cognition lies in the nature of the mental representation they involve. Sensory integration operates purely on sensory features (e.g., orientation, motion direction, color), processed by selective cells in sensory cortex *(Jazayeri and Movshon, 2006; Pouget et al., 2003)*. By contrast, higher-order processes – such as abstract categorization *(Cheadle et al., 2014; Wyart et al., 2015, 2012)* – or rule discovery *(Maheu et al., 2022)* – appear to require symbolic representations that are detached from the sensory features on which they are constructed. These symbolic representations enable cognitive transformations – such as compression *(Al Roumi et al., 2021)*, reversal *(Xie et al., 2022)*, reordering *(Albouy et al., 2022),* or the transfer of temporal structure across contexts *(Garvert et al., 2017; Lerousseau and Summerfield, 2024; Whittington et al., 2020)*. Our task formally tests this distinction through two competing hypotheses:

- (H_1) Sensory integration and rule discovery are independent and rely on different mental representations. Human predictions in the task will simply reflect the aggregate output of both processes.
- (H_2) Sensory integration can also operate on symbolic representations. Discovering the hidden rule will trigger the symbolic transformation of presented sensory features prior to their integration.

To investigate this possible bidirectional interplay between sensory integration and rule discovery (H_2), we asked participants to predict the final orientation of a sequence of ten stimuli sequentially flashed (**Fig. 1A**). The experiment comprised two types of blocks. In the control blocks (*stay* blocks – gray), the orientation of each stimulus was drawn from a normal distribution centered around a uniformly sampled mean. In the other blocks (*switch* blocks - purple), we introduced a hidden rule: the last five stimuli were drawn from a distribution shifted by 90° compared to the first five. In each block, after the presentation of complete sequences (training), participants were presented with different conditions – incomplete sequences consisting of 3, 5, 7 or 9 stimuli. They had to predict the orientation of the final (10th) sample – meaning a n+7 prediction for incomplete sequences of 3 stimuli, a n+5 prediction for incomplete sequences of 5 stimuli, etc. (**Fig. 1B**). Critically, in the n+7 and n+5 conditions of the switch blocks, discovering the hidden rule was the only way to accurately perform the task.

Contrary to H_1, we found that the amount of leak during sensory integration in stay blocks predicted whether participants could discover the hidden rule in switch blocks. Consistent with H_2, discovering the rule led participants to mentally rotate each presented stimulus by 90° in switch blocks before integrating them. Finally, we used targeted manipulations of recurrent neural networks (RNNs - *(Kleinman et al., 2021; Kurikawa, 2024; Lin et al., 2022)* to further establish a causal link between integration leak and rule discovery.

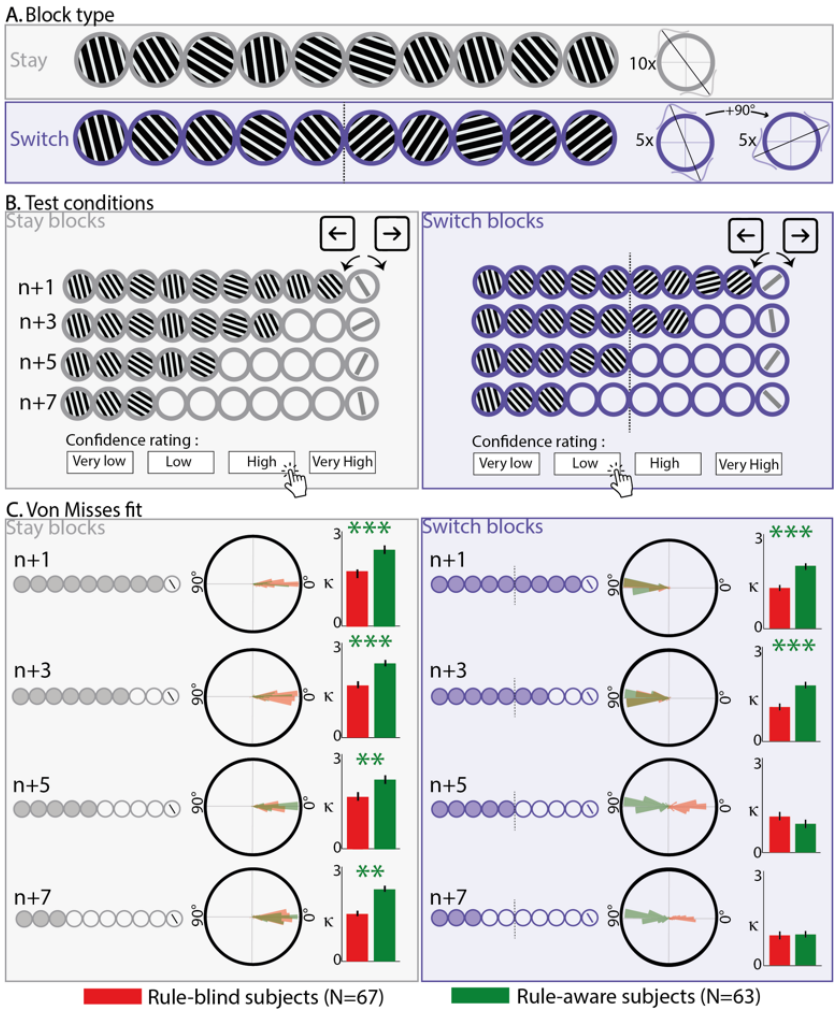


Figure 1 A. *Description of the two types of blocks. Each of the grating patches lasts 250ms. In stay blocks, sequences were composed of 10 items drawn from a normal distribution centered on a target orientation randomly picked for each sequence. For switch blocks, the 5 last stimuli of sequences were drawn from the 90° shifted distribution compared to the 5 first.* ***B****. For each block, after the presentation complete sequences, participants performed different conditions, i.e., incomplete sequences of 3/5/7 or 9 first stimuli. For each, they had to predict what would have been the orientation of the last (10th) item of this particular sequence by rotating a bar.* ***C.*** *We compared participants' predictions to the mean orientation before switch and fitted Von Misses distributions for each participant, condition and block. Polar plots display the histogram of the location (θ) of the Von Misses distribution for each participant, while bar plots display participants' average precision (κ). **: $p_s<0.01$ *** $p_s<0.001$ (Bonferroni corrected across conditions).*

Results

Prediction precision as a behavioral signature of rule discovery

We identified behavioral signatures diagnostic of latent rule discovery, to classify participants accordingly. Separately for each type of block (stay/switch) and each condition (n+1/n+3/n+5/n+7), we fitted the distribution of tilt between each participant's response (their predicted orientation for the 10th stimulus) and the mean orientation before the switch (see **Methods**). The fitted location (θ) and concentration (κ) of these distributions provided clear signatures of rule discovery (**Fig. 1C**). Indeed, for conditions *n+5* and *n+7* in switch blocks, the orientation of the 10th stimulus is shifted by an average of 90° relative to the mean of the pre-switch distribution. A participant who discovered the hidden rule should have a location θ of 90°, whereas a participant who did not should have a location θ close to zero, as they have only seen the first half of the sequence in these conditions. Fitted location for these conditions in switch blocks revealed a clear bimodal distribution around these two behaviors (**Fig. 1C**). We splited participants into two groups based on their location in these conditions, using an absolute location of **45°** as the threshold for classifying participants either as *rule-aware* (N=63) or *rule-blind* (N=67, **Fig. 1C**).

We then compared these two groups in terms of other aspects of behavior. The first step was to compare their response precision κ (see **Methods**). Rule-aware participants made more precise predictions than rule-blind participants in *stay* blocks – i.e., in the absence of hidden rule (concentration κ: $p_s<0.01$, Bonferroni corrected; **Fig. 1C**). Rule-aware participants also made more precise *n+1* and *n+3* predictions in *switch* blocks ($p_s<0.001$). This improvement in prediction precision in rule-aware participants was accompanied by an increase in their self-reported confidence relative to rule-blind participants (**Fig. S1**).

To ensure that the link between prediction precision in stay blocks and rule discovery in switch blocks was not due to metacognitive disinvolvement of participants who failed to discover the hidden rule and therefore disengage from the task – we replicated our analysis restricted to the data from the first block of participants who started the experiment with a stay block. Prediction precision (κ) in this first stay block – before those participants had ever experienced a switch block – was significantly higher for participants labelled as rule-aware based on their behavior in subsequent *switch* blocks (p=0.014).

Sensory integration leak prevents rule discovery

The increased prediction precision of rule-aware participants suggests that the quality of sensory integration affects rule discovery. To further characterize the nature of the interaction between sensory integration and rule discovery, we used a computational modeling approach *(Drugowitsch et al., 2016; Findling and Wyart, 2021).* First, we decomposed each participant's errors in stay blocks into two sources: computation errors, arising from suboptimal integration of successive stimuli, and response errors, reflecting suboptimal reporting of the participant's prediction of the 10th stimulus (Model 1; **Fig. 2A**; see **Methods**). In contrast to response errors, computation errors accumulate across stimuli, leading to a linearly increasing variance of errors with sequence length allowing for a robust separation of these two error sources (see recovery-based analyses **Supplemental Information**).

We first confirmed that our results were not due to general disengagement by rule-blind participants. Computation errors were significantly greater for the rule-blind participants than for rule-aware participants (Ranksum, p<0.001), whereas the two groups did not differ in terms of their response errors (p=0.35 **Fig.2B**). Furthermore, the correlation between the two sources of errors across participants was very modest (R^2=0.04). Both findings are inconsistent with a general difference in task engagement between rule-aware and rule-blind participants and instead indicate a selective increase in computation errors in rule-blind participants.

To better understand the specific sensory integration parameters which affect rule discovery in switch blocks, we estimated a more detailed model of participants' behavior in stay blocks (Model 2). It comprised five distinct sources of suboptimality (**Fig. 2C**): 1/ sensory bias (γ_{sen}) that distorts stimulus orientation toward diagonals *(Drugowitsch et al., 2016; Girshick et al., 2011),* 2/ random noise of standard deviation σ_{int} that corrupts the integration of stimuli, 3/ leak (α) in the integration process, 4/ random noise of standard deviation σ_{rep} that corrupts the reporting of the predicted orientation of the 10th stimulus, and 5/ reporting bias (γ_{rep}) that distorts the predicted orientation toward diagonals. Parameter recovery confirmed that all parameters were accurately dissociable and meaningfully interpretable (**Fig. S2**).

Fitting this model to participant's behavior in *stay* blocks revealed a selective difference in sensory integration between rule-aware and rule-blind participants: rule-aware participants showed lower integration leak α (0.32+/-0.05) than rule-blind participants (0.80+/-0.08; Ranksum p<0.001 Bonnferoni corrected). No other parameter significantly differed between the two groups of participants, though integration noise σ_{int} was marginally higher in rule-blind participants (rule-aware:0.93+/-0.05, rule-blind:1.4+/-0.2, p=0.054).

To confirm this relationship between integration leak and rule discovery, we applied a cross-validated logistic regression to predict each participant's label (rule-aware or rule-blind) using all five parameters of Model 2 fitted to their behavior in stay blocks only (see **Methods**). The logistic regression model predicted participants' label much better than chance, with an average binary accuracy of 74% (95% CI =[0.677, 0.799]). Importantly, only integration leak and integration noise had significant coefficients for predicting rule discovery ($\bar{\sigma}_{int}$ = 1.8 [0.94,2.6], $\bar{\alpha}$ = 1.5[0.87,2.3], **Fig. 2E**).

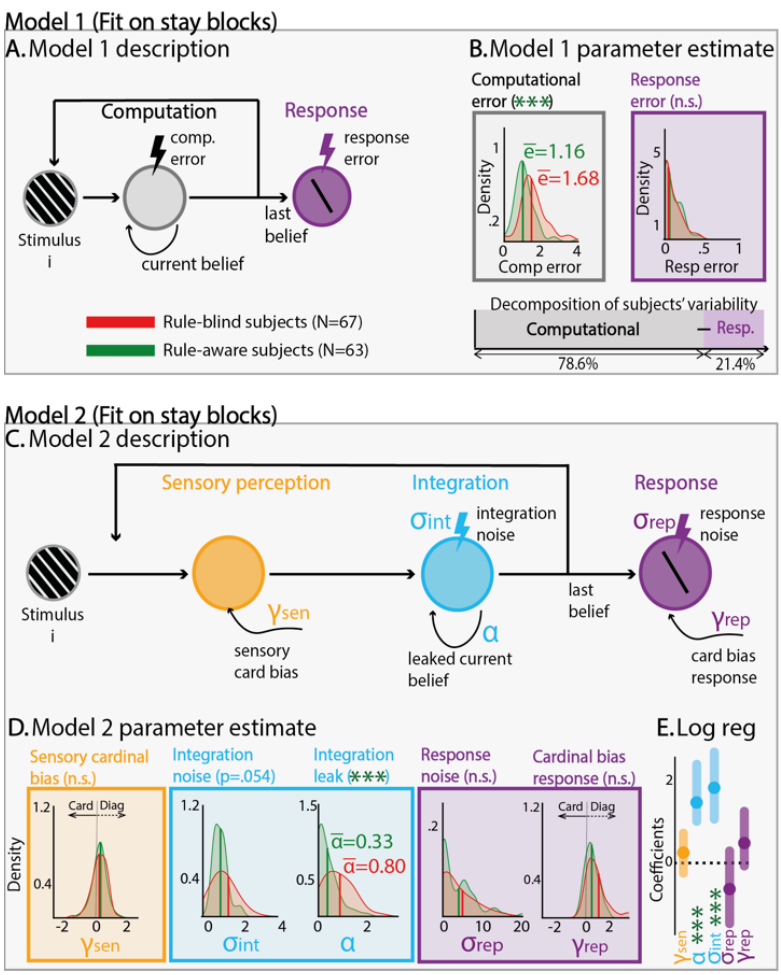


Figure 2 : A. *Description of Model 1 separating computational from response errors.* ***B.*** *Parameter fit for each participant on stay blocks and group comparison (distribution of parameters estimate are plotted with kernel density estimate, vertical bar indicates the average of the group. ***: p_s<0.001 (Bonferroni corrected over the 2 parameters) .* ***C.*** *Description of Model 2 with five suboptimalities parameters.* ***D.*** *Parameter fit on stay blocks and group comparison ***: p_s<0.001 (Bonferroni corrected over the 5 parameters)* ***E.*** *Multiple logistic regression to predict rule discovery from all 5 parameters. Dots represent coefficient estimates and bars 95% confidence intervals. ***: p_s<0.001 (Ranksum test Bonferroni corrected over the 5 parameters).*

Rule discovery affects sensory integration

Analysis of the stay blocks provided clear evidence for a "bottom-up" interaction between sensory integration and rule discovery: participants who integrated stimuli with less leak (and increased precision) in stay blocks were more likely to discover the hidden rule in switch blocks. To explore a "top-down" interaction of rule discovery on human integration strategy, we turned to the analysis of participants' behavior in switch blocks. We compared two hypotheses: Either participants that discovered the rule do not change their integration process but only add a correction offset to their response (H1; Model 3), or they flexibly apply mental rotation to each of the 5 first stimuli, prior to sensory integration (H2; Model 4).

We fitted those two models (**Fig. 3A***)* on switch data together with the baseline model without any offset (Model 2; **Fig. 2A**). Model comparison (see **Methods**) revealed the integration offset model to be the best fit for rule-aware participants (AIC: Model2=671, Model3=647, Model4=523; $p<0.001$, **Fig. 3A**). To check that the offset rule implemented this Model was correct (same offset for the five first stimuli and no offset for the five last), we fitted a complementary model with 9 offsets, one for each stimulus. This analysis confirmed our hypothesis, showing rule-aware participants had an exact representation of the switch rule and applied a 90° offset to the 5 first stimuli and no offset to the 5 last (**Fig. S3**). This analysis suggests that discovering the rule modified participant's sensory integration process, rather than simply accounting for the rule in their reporting process.

To further support this idea, we explored its predictable implications. Mentally rotating perceived orientations prior to integration should be associated with added imprecision. This should cause an increase of integration noise in switch blocks compared to stay blocks for rule-aware participants only. We compared integration noise during stay and switch blocks, and found that rule-aware participants had an increased integration noise in the switch compare to the stay blocks (σ_{int} switch=1.6+/-0.099, stay=0.93+/-0.049, Ranksum $p<0.001$) while it was not the case for rule-blind participants (σ_{int} switch=1.49+/-0.19, stay=1.35+/-0.17, $p=0.8$; **Fig. 3B**). Additionally, we reasoned that only participants who did not discover the rule should perceive switch blocks as more volatile than stay blocks. We tested this hypothesis by comparing the integration leak during the stay and the switch blocks for both groups, as a greater perceived volatility is captured by this specific parameter. We indeed found an increased leak in switch compared to stay blocks for rule-blind participants (α switch=1.5+/-0.13, stay=0.79+/-0.08; Ranksum $p<0.001$), up to the point where they almost ignored pre-switch stimuli, while rule learners did not adjust their leak revealing comparable perceived volatility (α switch=0.31+/-0.06, stay=0.32+/-0.05; $p=0.6$; **Fig. 3B**).

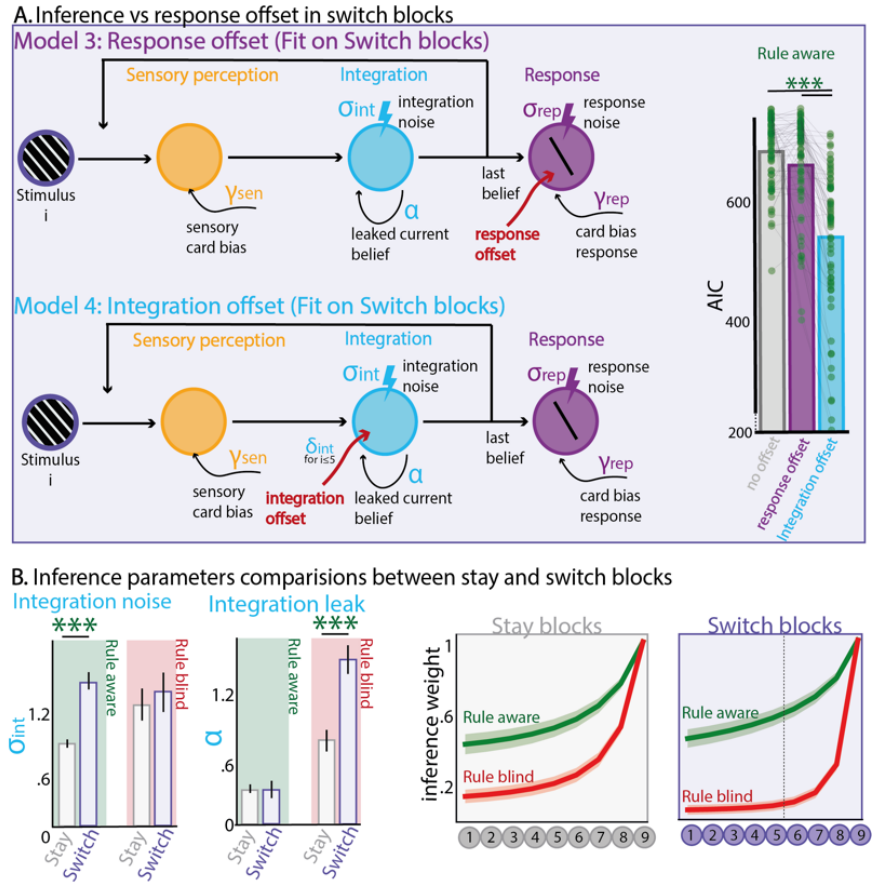


Figure 3 : A. Left. *Comparison of Models 3 and 4. Model 3 is similar to Model 2 (fig. 2) with an additional offset parameter applied to the response. Model 4 has instead an offset parameter applied to the five first stimuli of the sequence.* ***Right.*** *AIC comparison when fitted on participants' responses to switch blocks for Models 2 ('no offset'), 3 ('response offset'), and 4 ('integration offset').* ***B.*** *Comparison of the two integration parameters – noise* σ_{int} *and leak* α *– within participant, between stay and switch blocks, for rule-aware and rule-blind participants. Left panels represent the weight associated with each element of the sequence for the prediction based on the leak value fitted for each participant. ***: p<0.001 (Ranksum test Bonferroni corrected for two comparisons).*

Integration leak causaly prevent rule discovery in Recurrent neural Networks

Previous analyses showed that participants' recency bias — i.e., their integration leak — was associated with their capacity to uncover the hidden rule in switch blocks. To investigate a possible causal relation between sensory integration and rule discovery, we studied the behavior of recurrent neural networks (RNNs) on the same tasks. To mimic human-like sequential integration, we trained RNNs to predict each upcoming stimulus in the sequence, rather than only the last (10th) one. This approach allowed us to measure the network's surprise at each stimulus in the sequence (see **Methods**). Based on these surprise estimates, we defined a *Rule Learning Index (RLI)* as the network's 'surprise' to the switch (stimulus 6), relative to the average surprise across all other stimuli. This index captures the dynamic of the ability of the network to correctly predict the switch rule and provides a quantitative metric of rule discovery in RNNs. Throughout the training, we observed a rapid decline in surprise for all stimuli except the switch, corresponding to the learning of pure sensory integration. This was followed by a later decrease of the surprise evoked by the 6th stimulus, indicating rule discovery (**Fig. 4B**). This dynamic is revealed in a sharp initial rise in the RLI (sensory integration learning), followed by a slower decrease (rule learning) until convergence.

Next, we introduced selective perturbations to replicate the suboptimalities observed in human behavior. We created five variants, each implementing one of the suboptimalities described by the five parameters of Model 2 (**Fig. 2**). The sensory bias toward diagonals (γ_{sen}) was implemented at the input level of the network. Integration noise (σ_{int}) was obtained from noisy connections in the hidden layer. Integration leak (α) was obtained by sparsifying the hidden layer, forcing a certain percentage of weights to be fixed at zero. Finally, the two response suboptimalities $\left(\gamma_{rep}\ \&\ \sigma_{rep}\right)$ were applied at the output level of the network. For each variant, we implemented five levels of suboptimalities (from low to high) and trained

20 networks per level. We simulated the behavior of trained RNNs on the task and verified that each manipulation selectively affected only the intended parameter by fitting Model 2 to the RNN's behavior in the stay blocks (as we did with human data). Each manipulation selectively affected the targeted parameter, with the exception of the sensory bias toward diagonals, that also affected the response cardinal bias (**Fig. 4C**).

We used the *RLI* to assess how each suboptimality causally affects the learning trajectory in switch trials and showed that, in line with the findings from human data, only the two integration parameters (σ_{int} & α) modified RLI dynamics. More precisely, while adding integration noise (σ_{int}) via noisier recurrent connections impaired both sensory integration and rule discovery, increasing integration leak (α) specifically delayed discovery of the switch rule without altering sensory integration (**Fig. 4C**). We confirmed that this relationship was not limited to the particular case of a single switch in the middle of the sequence by replicating this analysis using sequences of the form AABBAABBAA (with multiple switches every two stimuli) (**Fig. S4**).

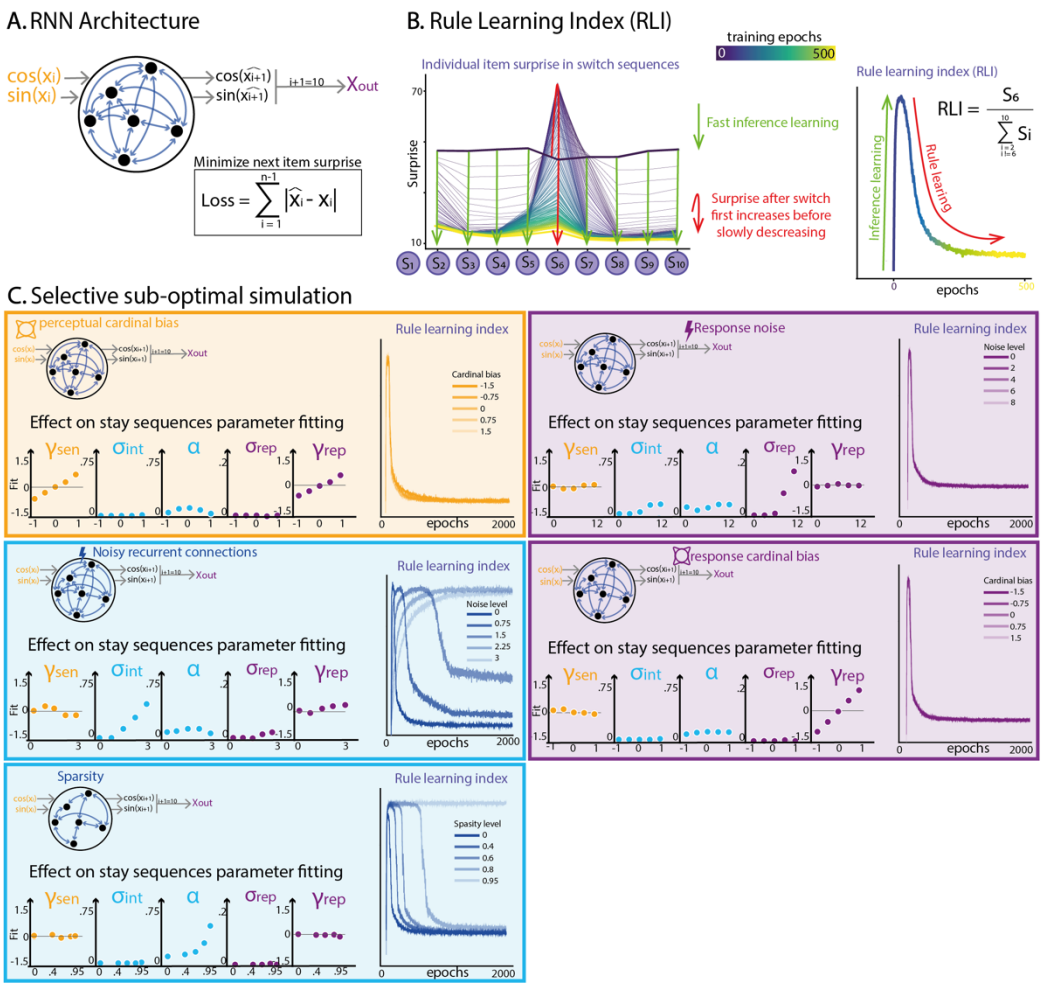


Figure 4 : A. *Architecture of the Recurrent Neural Network (RNN). The RNN receives as input the current stimulus orientation (x_i), and is trained to predict the next orientation (x_{i+1}). The difference between the predicted and actual orientation defines the 'surprise'.* ***B. Left:*** *Surprise for each stimulus (2-10) in the switch blocks as a function of training epochs (color-coded from purple to yellow). Green arrows highlight the rapid decrease in surprise for stimuli 2/3/4/5 and 7/8/9/10 corresponding to the learning of sensory integration predictions. The red arrow highlights the slower decrease in surprise following the switch, corresponding to the learning of the rule.* ***Right:*** *This learning dynamic is summarized by the Rule Learning Index (RLI), defined as the ratio of surprises related to rule learning (stimulus 6) and sensory integration (stimuli 2-to-5 and 7-to-10).* ***C.*** *Five variants of the RNN designed to*

implement each of the suboptimalities defined in Model 2, with five levels of suboptimalities (from low to high) for each . 1. Perceptual cardinal bias applied to each orientation xi. 2. Integration noise modeled via noisy computation in the recurrent layer. 3. Integration leak sparsifying to the recurrent layer. 4&5. Response-related parameters were applied at the output level (Xout). ***Bottom graphs on each panel.*** *After training we fitted Model 2 to each RNN's predictions on stay blocks to estimate how each manipulation impacted each of the parameters.* ***Right graph on each panel.*** *For each level of suboptimality, we estimated the RLI to evaluate how it affected the learning dynamic.*

Discussion

To make sense of sequences of stimuli, humans rely on multiple learning abilities including temporal interval prediction *(Janata et al., 2012; Nobre and van Ede, 2018; Wollman and Morillon, 2018; Zalta et al., 2020)*, sensory integration *(Bogacz, 2007; Morillon et al., 2014; Summerfield and Egner, 2009; Wald and Wolfowitz, 1948)*, categorization based on arbitrary mappings of sensory features *(Cheadle et al., 2014; Freedman and Assad, 2016; Wyart et al., 2012)*, learning statistical dependencies between stimuli *(Benjamin et al., 2024, 2023; Henin et al., 2021; Saffran et al., 1996; Schapiro et al., 2016, 2013)*, and the learning of hidden rules that describe the patterns of relations between successive stimuli *(Al Roumi et al., 2023, 2021; Albouy et al., 2022; Quilty-Dunn et al., 2022; Xie et al., 2022)*. In a majority studies, these different learning abilities are assumed, either implicitly or explicitly, to rely on different cognitive and neural mechanisms (for a counterpoint, see *Fiser & Lengyel, 2022*), but they have rarely been studied conjointly. A few recent studies have begun to address this issue, by investigating the impact of temporal prediction on statistical learning *(Gómez Varela et al., 2024)*, or by asking human participants to classify sequences of stimuli drawn either from deterministic rules or from probabilistic distributions *(Maheu et al., 2020).* The present study aims at understanding how these different learning abilities interact to form a cognitive system that is able to make sense of sensory events unfolding in time. Our results revealed a strong two-way interplay between sensory integration and rule discovery: (1) a bottleneck exerted by the recency bias (leak) of sensory integration on the probability of discovering the hidden rule, and (2) the application of the hidden rule to incoming stimuli prior to their integration.

First, the results show that two previously described suboptimalities in sensory integration – integration leak and integration noise *(Drugowitsch et al., 2016; Rahnev and Denison, 2018; Waskom and Kiani, 2018; Wyart and Koechlin, 2016)* – when measured in sequences which do not contain a hidden rule, predict the likelihood of discovering a hidden rule in independent sequences (**Fig. 2E**). Among these two suboptimalities, the amount of integration leak (which reflects the timescale of sensory integration) showed the strongest difference between participants who did discover the hidden rule and those who did not. Note that the hidden rule in switch sequences triggered not a gradual but a brutal 90° shift in mean stimulus orientation between the 5th and 6th stimuli, which therefore did not require a long timescale to be detected. Participants who did not find the hidden rule in switch blocks showed moderate, not extreme integration leak in stay blocks.

To disentangle the effects of these two integration suboptimalities on rule discovery, we applied selective causal perturbations to RNNs trained to perform the same task as human participants. This effort is part of a growing body of research that leverage these flexible artificial systems to understand cognition, not by fitting their free parameters (connection weights) to behavior *(Eckstein et al., 2023; Peterson et al., 2021)*, but by analyzing the structural and functional configurations which, through constrained optimization,

result in the same suboptimalities as biological brains *(Findling and Wyart, 2024; Flesch et al., 2022; Molano-Mazón et al., 2023)*. Sparsifying recurrent connections produced a graded increase in integration leak during stay blocks without affecting noise, while adding computational noise in recurrent units *(Findling and Wyart, 2024, 2021)* increased integration noise independently of leak. In switch blocks, these manipulations revealed a double dissociation: increasing sparsity impaired rule discovery without disrupting sensory integration (albeit making it leakier), whereas increasing noise impaired both processes. This distinction was captured in RNN surprise dynamics during training: networks first reduced surprise to all stimuli except the 6th (indicating sensory integration), then gradually to the 6th (indicating rule discovery). Only integration leak delayed the latter without altering the former. Control perturbations targeting sensory or response biases had no effect on either process. Finally, the link between leak and delayed rule discovery generalized to sequences with a different hidden rule structure (with 90° shifts every two stimuli), confirming the robustness of this causal relationship.

While integration leak was reliably and consistently tied to poorer rule discovery abilities both in humans and RNNs, our results also show that discovering the hidden rule modified how human participants integrate stimuli in switch blocks. Specifically, human participants applied the hidden rule to individual stimuli prior to integrating them in order to predict the final (10th) stimulus. This finding is directly contradicting dominant theories and computational models of sensory integration, which typically postulate that integration occurs within the 'native sensory space' defined by stimulus features (e.g., orientation) *(Deneve et al., 2001; Ermentrout, 1998; Esnaola-Acebes et al., 2021; Jazayeri and Movshon, 2006; Pouget et al., 2003)*. If sensory integration was encapsulated in this way, rule discovery would result in a post-hoc shift applied to the integrated result. This is not what we observed: participants' behavior was best explained by a model in which stimuli were mentally rotated according to the hidden rule *prior* to integration. Such mental manipulations of stimuli have already been described in rule learning tasks, notably in the context of the learning of temporal rules in stimulus sequences, such as the reordering *(Albouy et al., 2022; Xie et al., 2022)*, mapping *(Garvert et al., 2017)* or compression of stimuli *(Planton et al., 2021)*. These findings are much sparser in the context of sensory integration. However, tasks requiring the categorization of stimulus sequences along nonlinear functions of stimulus features (e.g., a cardinal/diagonal rule for stimulus orientation) have identified dissociated neural correlates of the sensory processing of individual stimuli and the integration of their associated category evidence *(Cheadle et al., 2014; Wyart et al., 2015, 2012)*. Our findings extend this work by showing that sensory integration itself can operate over transformed, symbolic representations—suggesting that the cognitive system flexibly reconfigures the format of information prior to integration when higher-order structure is detected.

Our study is not the first to investigate the impact of hidden rule discovery on decision making. However, the hidden rules used in previous studies typically did not interact with the processing of presented stimuli. For example, the hidden rule used in Schuck et al *(2015)* afforded bypassing entirely the processing of the pattern of dots to focus solely on their color to decide. This bypass prevented testing for the interplay between stimulus processing and rule discovery. By contrast, in our task, a leak in sensory integration directly prevented participants from discovering the hidden rule, an objective limitation in constructing accurate predictions of upcoming stimuli. RNN perturbations showed that integration leak emerged only for extremely sparse networks, resulting in a substantial decrease in the number of computations performed. Indeed, RNNs with only 5% of active connections showed an integration leak comparable to human participants. The integration leak measured in human participants could therefore partly stem from a cost-benefit optimization,

which trades the benefits of efficient sensory integration and rule discovery against the cost of more dense computations. Such cost-benefit arbitration has been postulated to explain other suboptimalities in decision-making such as integration noise, which could arise from the flexibility of associative circuits that can map arbitrary functions of sensory features (here, the integration of stimulus orientations with rule-predicted shifts) onto actions *(Findling and Wyart, 2024, 2021)*. Indeed, the attentional selection and pooling of sensory signals according to a current context *(Mante et al., 2013; Pestilli et al., 2011)* by low-rank associative neural circuits has been theorized to be critical for decision accuracy *(Dubreuil et al., 2022)*.

Together, these results revealed a profound bidirectional interplay between sensory integration and rule discovery. It highlights the necessity to conjointly study those cognitive abilities to obtain a clear picture of how humans are able to build a rich mental representation of uncertain sensory environments.

Materials and Methods

Participants

130 UK based adults (>18yo) were recruited online via the prolific platform and performed two sessions the task on consecutive days (52 female, mean age = 41,5). They were paid 8£ for the first and 9£ for the second session. Participants gave informed consent and the study was approved by Inserm Ethical Review Committee (IRB00003888 on 13/11/2018).

Experimental paradigm and stimuli

The experiment consists two types of blocks: in *stay* blocks, each stimulus of the sequence is a grating patch whose orientation is drawn from a normal distribution ($\sigma = 15°$). In *switch* blocks, we added a 90° rotation from the middle of the sequences. Importantly the means of the distributions were randomly chosen for each trial so that the abstract +90° rule must be learnt. Each sequence was sequentially presented at 2Hz. The block type was indicated by the color of the circle surrounding the stimuli. Block order was balanced across participant. In each block, after the presentation of 6 full training sequences, participants were presented 20 with incomplete sequences, stopping after 3, 5, 7 or 9 stimuli (5 of each), and had to predict the orientation of the last (10th) stimuli by rotating a bar, and estimate their confidence level. 4 full sequences were randomly placed between the incomplete test sequence to consolidate participant learning. After each answer, they saw colored feedback indicating how far they are from the real orientation. They performed a total of 400 test trials over the two sessions.

Preregistration

The procedure, as well as the primary outcome of this analysis (bidirectional interactions between sensory integration and rule discovery), has been pre-registered prior to data collection. The relevant details can be found here: https://aspredicted.org/wy43v.pdf

Von Misses distributions fit

We performed model free analysis by fitting von Misses distributions of the difference between participants' answers and the average orientation of the five first stimuli (only three first for n+7 conditions) of the sequence for each condition separately. Von Misses distribution have two parameters : location (θ) which indicates the average direction of the distribution, and the concentration (κ), which reflects variability of the response.

Model design and fit

We performed model-based analyses with several models :

Model 1 : This model performs integration of evidence with two sources of uncertainties: computational and response errors. The first is modelled by a random error for each orientation drawn from a centered normal distribution with a variance σ_{inf}. The latter is modelled by a random error on the final response with variance σ_{rep}. To estimate the impact of each error source on the overall variance, we fitted this model to estimate σ_{inf} & σ_{rep} for each participant and simulated the response with σ_{inf} only, σ_{rep} only or both. We computed the average error for those simulations. The variance attributed to σ_{inf} is the ratio between the average error of the simulation with σ_{inf} and the average error with both sources.

Model 2 : This second model comprises 5 parameters. Sensory cardinal bias (γ_{sen}), integration noise (σ_{int}), integration leak (α), response noise (σ_{rep}) and response cardinal bias (γ_{rep}). σ_{int} and σ_{rep} were modeled like in Model 1. For the sensory cardinal bias, each orientation of the sequence was modified through the function $f(x,g) = x + \frac{\sum_n sin(4nx) I_n(g)/n}{2I_0(g)}$, with x the orientation, g the bias strength and In the Bessel function of order n. For the response cardinal bias, the same function was applied on the response. The leak was modeled with an exponential decrease affecting the weight attributed to each stimulus in the participant response *(Ossmy et al., 2013; Usher and McClelland, 2001; Wyart et al., 2015)*.

Model 1 and 2 were fitted to stay blocks to estimate participants' parameters when performing a sensory integration task.

Model 3 : Model 3 is equivalent to Model 2, but we added a parameter modeling a possible fixed offset that participant could apply to all answers.

Model 4 : Model 4 is equivalent to Model 2 but with a fixed offset applied to the five first orientation of each sequence.

Model 3 and 4 have been fitted to switch sequence to disentangle between integration and response offset in case of rule learning (**Fig. 3A**)

Model 5 : Variation of Model 4 presented in SI, were each of the 9 stimuli of the sequence can have a different offset.

To estimate parameters, we fitted each model using BADS toolbox in MATLAB with 10000 bootstrap resamples to minimize the likelihood estimation an find best fitting parameters in each case. Model 1 and 2 were fitted on data from stay blocks only while Models 3,4 & 5 were fitted on data from switch blocks only. We assessed for statistical significance between parameters models using ranksum test given the non-normality some parameter distributions and applied Bonferroni correction.

Model recovery procedure: To ensure model validity, we performed parameters recovery on Model 2 by simulating 130 fake participants with parameter values pseudo-randomly drawn, ensuring that parameters distributions were uncorrelated (<1%). We fitted simulated stay blocks with Model 2 and compared fitted with generative parameter values (**Fig. S2**). All correlations between generative and fitted parameters exceeded 63% (γ_{sen}: 63%, σ_{inf}: 77%, α: 88%, σ_{rep}: 74%, γ_{rep} = 65%) showing a high reliability of fitted parameters estimate. Moreover, fitted parameters were uncorrelated to each other (<10%) ensuring that all parameters are correctly dissociable and do not suffer from tradeoff during the fitting procedure. To further prove model capacity in capturing participants behavior, we simulated each participant based on its own best

fitting parameters. We applied to this simulated dataset the same Von Misses procedure applied to the real data and compared location (θ) and concentration (κ) parameters obtained. We found that the simulated participants faithfully replicated human data, with high correlation between simulated and real behavioral metrics. The same analysis was performed on the switch blocks with Model 4. All those controls ensure the high reliability of our models in capturing humans' data.

RNN

We first designed a simple RNN composed of a 64x64 recurrent layer. The input and output orientations are coded by a two-dimensional array with cos and sin. Importantly, the loss of the RNN was chosen to mimic human like computation. Indeed, when given an input, the networks predicts the orientation of the next stimulus. The absolute value of the difference between the prediction and the actual orientation is called the 'surprise' and the loss of the sequence is the sum of the surprises in the sequence. That way, networks are trained with a human-like strategy.

We then implemented each suboptimalities in the RNN. The sensory cardinal bias was introduced by applying a cardinal bias for each stimulus of the sequence before the input (see Model design for the implementation of this bias). We used five cardinal bias values : [-1, -0.5, 0, 0.5, 1]. To implement the leak we made the recurrent layer sparse by fixing a given proportion of the connections to zero. This proportion was fixed to 0, 40%, 60%, 80% or 95%. To implement integration noise, we added random noise in the computation of the recurrent layer activity from a centered normal distribution with a std of 0, 0.75, 1.5, 2.25 or 3. The two response parameters did not modify the RNN per se, but were applied on Xout : either random noise or cardinal bias. To check that those modifications were specifically affecting only one parameter of the sensory learning mechanism, we trained 20 randomly initialized RNNs on 2000 epochs of 32 complete stays sequences, for each parameter value.

We then simulated the behavior after training on an occurrence of our experiment (for comparison purposes between the RNNs, we always used the same 20 tasks occurrences that correspond to the task presented to the first 20 participants). We could therefore fit the simulation results of the stay sequences with Model 2 and plot the impact of the manipulation on the fitted parameters, averaged across the 20 RNNs (**Fig. 4C**). We explored the impact of these modifications by plotting a summary metric we called *Rule Learning Index* (RLI) capturing the dynamic of the integration and rule learning though training. This metric was computed as the ratio of the surprised elicited by the 6th stimulus and the average surprised elicited by all other stimuli (**Fig. 4B**).

Acknowledgments

This work is supported by a FMR postdoctoral grant Espoir de la recherche awarded to L.B., a grant from the French National Research Agency (MONODEC, ANR-23-CE37-0028) awarded to V.W. as well as an European grant (ERC, SPEEDY, ERC-CoG-101043344) awarded to B.M.

References

Al Roumi F, Marti S, Wang L, Amalric M, Dehaene S. 2021. Mental compression of spatial sequences in human working memory using numerical and geometrical primitives. *Neuron* **109**:2627-2639.e4. doi:10.1016/j.neuron.2021.06.009

Al Roumi F, Planton S, Wang L, Dehaene S. 2023. Brain-imaging evidence for compression of binary sound sequences in human memory. *eLife* **12**:e84376. doi:10.7554/eLife.84376

Albouy P, Martinez-Moreno ZE, Hoyer RS, Zatorre RJ, Baillet S. 2022. Supramodality of neural entrainment: Rhythmic visual stimulation causally enhances auditory working memory performance. *Science Advances* **8**:eabj9782. doi:10.1126/sciadv.abj9782

Beck JM, Ma WJ, Pitkow X, Latham PE, Pouget A. 2012. Not Noisy, Just Wrong: The Role of Suboptimal Inference in Behavioral Variability. *Neuron* **74**:30–39. doi:10.1016/j.neuron.2012.03.016

Benjamin L, Fló A, Al Roumi F, Dehaene-Lambertz G. 2023. Humans parsimoniously represent auditory sequences by pruning and completing the underlying network structure. *eLife* **12**:e86430. doi:10.7554/eLife.86430

Benjamin L, Sablé-Meyer M, Fló A, Dehaene-Lambertz G, Roumi FA. 2024. Long-Horizon Associative Learning Explains Human Sensitivity to Statistical and Network Structures in Auditory Sequences. *J Neurosci* **44**:e1369232024. doi:10.1523/JNEUROSCI.1369-23.2024

Bogacz R. 2007. Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences* **11**:118–125. doi:10.1016/j.tics.2006.12.006

Cheadle S, Wyart V, Tsetsos K, Myers N, de Gardelle V, Herce Castañón S, Summerfield C. 2014. Adaptive Gain Control during Human Perceptual Choice. *Neuron* **81**:1429–1441. doi:10.1016/j.neuron.2014.01.020

Clegg BA, DiGirolamo GJ, Keele SW. 1998. Sequence learning. *Trends in Cognitive Sciences* **2**:275–281. doi:10.1016/S1364-6613(98)01202-9

Dehaene S, Al Roumi F, Lakretz Y, Planton S, Sablé-Meyer M. 2022. Symbols and mental programs: a hypothesis about human singularity. *Trends in Cognitive Sciences* **26**:751–766. doi:10.1016/j.tics.2022.06.010

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:10.1016/j.neuron.2015.09.019

Deneve S, Latham PE, Pouget A. 2001. Efficient computation and cue integration with noisy population codes. *Nat Neurosci* **4**:826–831. doi:10.1038/90541

Drugowitsch J, Wyart V, Devauchelle A-D, Koechlin E. 2016. Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality. *Neuron* **92**:1398–1411. doi:10.1016/j.neuron.2016.11.005

Dubreuil A, Valente A, Beiran M, Mastrogiuseppe F, Ostojic S. 2022. The role of population structure in computations through neural dynamics. *Nat Neurosci* **25**:783–794. doi:10.1038/s41593-022-01088-4

Eckstein MK, Summerfield C, Daw ND, Miller KJ. 2023. Predictive and Interpretable: Combining Artificial Neural Networks and Classic Cognitive Models to Understand Human Learning and Decision Making. doi:10.1101/2023.05.17.541226

Ermentrout B. 1998. Neural networks as spatio-temporal pattern-forming systems. *Rep Prog Phys* **61**:353. doi:10.1088/0034-4885/61/4/002

Esnaola-Acebes JM, Roxin A, Wimmer K. 2021. Bump attractor dynamics underlying stimulus integration in perceptual estimation tasks. doi:10.1101/2021.03.15.434192

Findling C, Wyart V. 2024. Computation noise promotes zero-shot adaptation to uncertainty during decision-making in artificial neural networks. *Science Advances* **10**:eadl3931. doi:10.1126/sciadv.adl3931

Findling C, Wyart V. 2021. Computation noise in human learning and decision-making: origin, impact, function. *Current Opinion in Behavioral Sciences* **38**:124–132. doi:10.1016/j.cobeha.2021.02.018

Fiser J, Lengyel G. 2022. Statistical Learning in Vision. *Annu Rev Vis Sci* **8**:265–290. doi:10.1146/annurev-vision-100720-103343

Flesch T, Juechems K, Dumbalska T, Saxe A, Summerfield C. 2022. Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron* **110**:1258-1270.e11. doi:10.1016/j.neuron.2022.01.005

Freedman DJ, Assad JA. 2016. Neuronal Mechanisms of Visual Categorization: An Abstract View on Decision Making. *Annu Rev Neurosci* **39**:129–147. doi:10.1146/annurev-neuro-071714-033919

Garvert MM, Dolan RJ, Behrens TEJ. 2017. A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife* **6**:1–20. doi:10.7554/eLife.17086

Girshick AR, Landy MS, Simoncelli EP. 2011. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat Neurosci* **14**:926–932. doi:10.1038/nn.2831

Gómez Varela I, Orpella J, Poeppel D, Ripolles P, Assaneo MF. 2024. Syllabic rhythm and prior linguistic knowledge interact with individual differences to modulate phonological statistical learning. *Cognition* **245**:105737. doi:10.1016/j.cognition.2024.105737

Henin S, Turk-Browne NB, Friedman D, Liu A, Dugan P, Flinker A, Doyle W, Devinsky O, Melloni L. 2021. Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances* **7**:1–13. doi:10.1126/sciadv.abc4530

Janata P, Tomic ST, Haberman JM. 2012. Sensorimotor coupling in music and the psychology of the groove. *J Exp Psychol Gen* **141**:54–75. doi:10.1037/a0024208

Jazayeri M, Movshon JA. 2006. Optimal representation of sensory information by neural populations. *Nat Neurosci* **9**:690–696. doi:10.1038/nn1691

Kleinman M, Chandrasekaran C, Kao J. 2021. A mechanistic multi-area recurrent network model of decision-makingAdvances in Neural Information Processing Systems. Curran Associates, Inc. pp. 23152–23165.

Kurikawa T. 2024. Different timescales of neural activities introduce different representations of task-relevant information. doi:10.1101/2024.07.23.604720

Lerousseau JP, Summerfield C. 2024. Space as a Scaffold for Rotational Generalisation of Abstract Concepts. *eLife* **13**. doi:10.7554/eLife.93636.1

Lin B, Bouneffouf D, Cecchi G. 2022. Predicting human decision making in psychological tasks with recurrent neural networks. *PLoS ONE* **17**:e0267907. doi:10.1371/journal.pone.0267907

Ma WJ, Kording K, Goldreich D. 2023. Bayesian models of perception and action: an introduction. London, England: The MIT Press.

Maheu M, Meyniel F, Dehaene S. 2022. Rational arbitration between statistics and rules in human sequence processing. *Nat Hum Behav* **6**:1087–1103. doi:10.1038/s41562-021-01259-6

Maheu M, Meyniel F, Dehaene S. 2020. Rational arbitration between statistics and rules in human sequence learning. doi:10.1101/2020.02.06.937706

Mante V, Sussillo D, Shenoy KV, Newsome WT. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**:78–84. doi:10.1038/nature12742

Molano-Mazón M, Shao Y, Duque D, Yang GR, Ostojic S, de la Rocha J. 2023. Recurrent networks endowed with structural priors explain suboptimal animal behavior. *Curr Biol* **33**:622-638.e7. doi:10.1016/j.cub.2022.12.044

Morillon B, Schroeder CE, Wyart V. 2014. Motor contributions to the temporal precision of auditory attention. *Nat Commun* **5**:5255. doi:10.1038/ncomms6255

Nobre AC, van Ede F. 2018. Anticipated moments: temporal structure in attention. *Nat Rev Neurosci* **19**:34–48. doi:10.1038/nrn.2017.141

Ossmy O, Moran R, Pfeffer T, Tsetsos K, Usher M, Donner TH. 2013. The timescale of perceptual evidence integration can be adapted to the environment. *Curr Biol* **23**:981–986. doi:10.1016/j.cub.2013.04.039

Pestilli F, Carrasco M, Heeger DJ, Gardner JL. 2011. Attentional Enhancement via Selection and Pooling of Early Sensory Responses in Human Visual Cortex. *Neuron* **72**:832–846. doi:10.1016/j.neuron.2011.09.025

Peterson JC, Bourgin DD, Agrawal M, Reichman D, Griffiths TL. 2021. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* **372**:1209–1214. doi:10.1126/science.abe2629

Planton S, van Kerkoerle T, Abbih L, Maheu M, Meyniel F, Sigman M, Wang L, Figueira S, Romano S, Dehaene S. 2021. A theory of memory for binary sequences: Evidence for a mental compression algorithm in humans. *PLoS Computational Biology.* doi:10.1371/journal.pcbi.1008598

Pouget A, Dayan P, Zemel RS. 2003. INFERENCE AND COMPUTATION WITH POPULATION CODES. *Annual Review of Neuroscience* **26**:381–410. doi:10.1146/annurev.neuro.26.041002.131112

Quilty-Dunn J, Porot N, Mandelbaum E. 2022. The Best Game in Town: The Re-Emergence of the Language of Thought Hypothesis Across the Cognitive Sciences. *Behavioral and Brain Sciences* 1–55. doi:10.1017/S0140525X22002849

Rahnev D, Denison RN. 2018. Suboptimality in perceptual decision making. *Behav Brain Sci* **41**:e223. doi:10.1017/S0140525X18000936

Ratcliff R, McKoon G. 2008. The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Comput* **20**:873–922. doi:10.1162/neco.2008.12-06-420

Saffran JR, Aslin RN, Newport EL. 1996. Statistical Learning by 8-Month-Old Infants. *Science* **274**:1926–1928.

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

Schapiro AC, Turk-Browne NB, Norman KA, Botvinick MM. 2016. Statistical learning of temporal community structure in the hippocampus. *Hippocampus* **26**:3–8. doi:10.1002/hipo.22523

Schuck NW, Gaschler R, Wenke D, Heinzle J, Frensch PA, Haynes J-D, Reverberi C. 2015. Medial Prefrontal Cortex Predicts Internally Driven Strategy Shifts. *Neuron* **86**:331–340. doi:10.1016/j.neuron.2015.03.015

Summerfield C, Egner T. 2009. Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences* **13**:403–409. doi:10.1016/j.tics.2009.06.003

Usher M, McClelland JL. 2001. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol Rev* **108**:550–592. doi:10.1037/0033-295x.108.3.550

Wald A, Wolfowitz J. 1948. Optimum Character of the Sequential Probability Ratio Test. *The Annals of Mathematical Statistics* **19**:326–339. doi:10.1214/aoms/1177730197

Waskom ML, Kiani R. 2018. Decision Making through Integration of Sensory Evidence at Prolonged Timescales. *Current Biology* **28**:3850-3856.e9. doi:10.1016/j.cub.2018.10.021

Whittington JCR, Muller TH, Mark S, Chen G, Barry C, Burgess N, Behrens TEJ. 2020. The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell* **183**:1249-1263.e23. doi:10.1016/j.cell.2020.10.024

Wollman I, Morillon B. 2018. Organizational principles of multidimensional predictions in human auditory attention. *Sci Rep* **8**:13466. doi:10.1038/s41598-018-31878-5

Wyart V, de Gardelle V, Scholl J, Summerfield C. 2012. Rhythmic Fluctuations in Evidence Accumulation during Decision Making in the Human Brain. *Neuron* **76**:847–858. doi:10.1016/j.neuron.2012.09.015

Wyart V, Koechlin E. 2016. Choice variability and suboptimality in uncertain environments. *Current Opinion in Behavioral Sciences*, Computational modeling **11**:109–115. doi:10.1016/j.cobeha.2016.07.003

Wyart V, Myers NE, Summerfield C. 2015. Neural Mechanisms of Human Perceptual Choice Under Focused and Divided Attention. *J Neurosci* **35**:3485–3498. doi:10.1523/JNEUROSCI.3276-14.2015

Xie Y, Hu P, Li J, Chen J, Song W, Wang X-J, Yang T, Dehaene S, Tang S, Min B, Wang L. 2022. Geometry of sequence working memory in macaque prefrontal cortex. *Science* **375**:632–639. doi:10.1126/science.abm0204

Zalta A, Petkoski S, Morillon B. 2020. Natural rhythms of periodic temporal attention. *Nat Commun* **11**:1051. doi:10.1038/s41467-020-14888-8

Supplementary Information

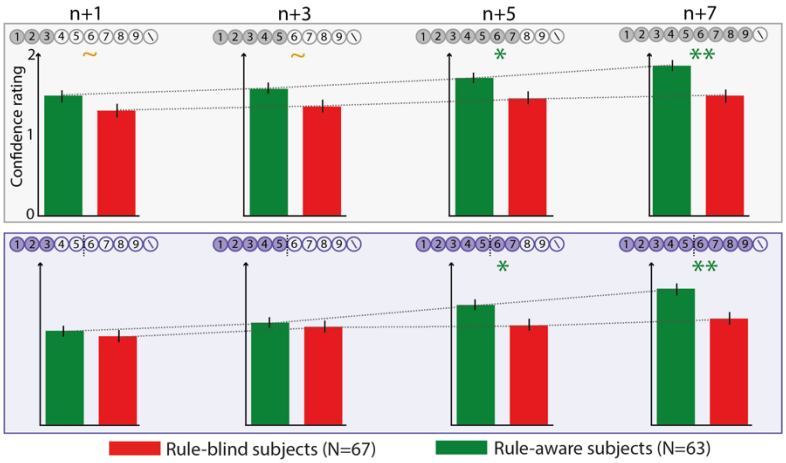


Figure S1 *: Confidence judgments for each condition, splited between participants who found the rule (green) and participants who did not (red). Congruently with their small integration leak, participants who found the rule increase their confidence with sequence length (they integrate more stimuli).*

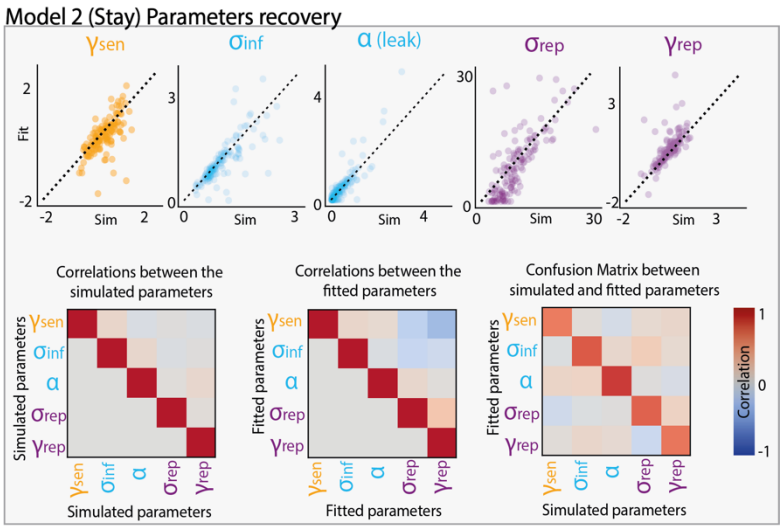
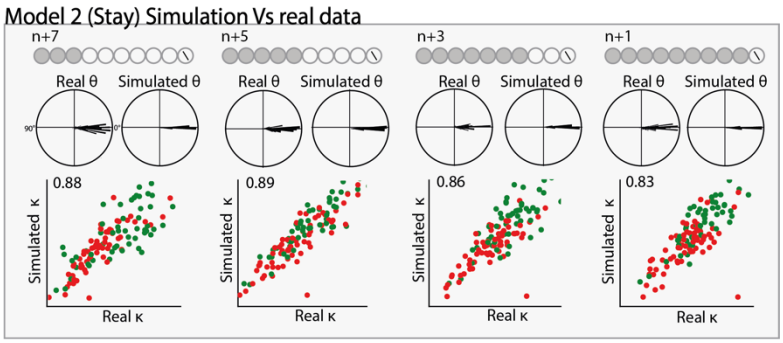
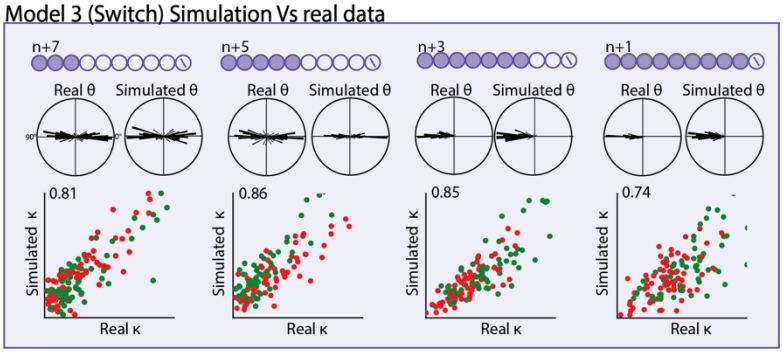


Figure S2 *: Validation of the different model. For Model 2 on stay sequences, we performed a parameter recovery and showed that all parameters could be correctly recovered. Moreover, the confusion matrix shows limited confusion between the different fitted parameters. For each participant fit, we simulated the result with Model 2 and compared the results of this simulation with the real data by comparing the von Misses parameters fits. We performed the same simulation vs real data comparison for Model 4 on switch sequences. Green dots represent participants who found the rule and red dots participants who did not.*





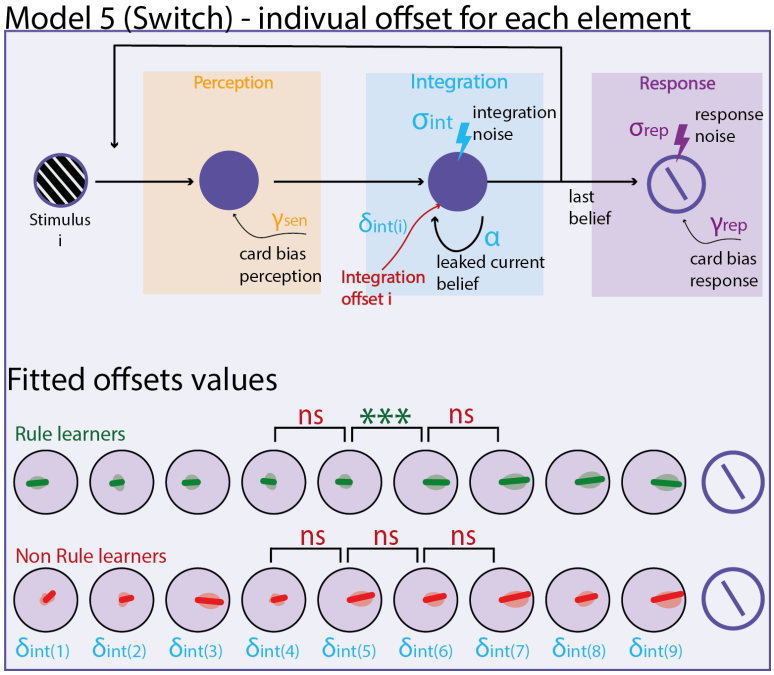


Figure S3 *: This model extends Model 4 but with an integration offset for each stimulus of the sequence. This new model confirms the validity of model 4 as all offsets of the first five stimuli are the same (~90°) with a clear switch between stimuli 5 and 6. Participants that didn't learn the rule, did not apply any offset to any stimulus of the sequence.*

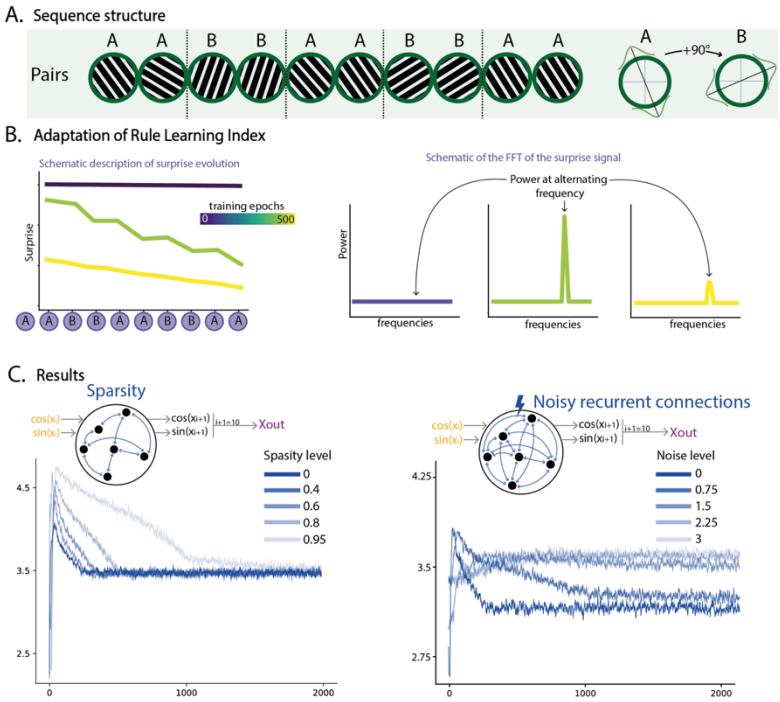


Figure S4 *: replication of the effect of sparsity (leak) and noise in recurrent connection (integration noise) on the effect of rule discovery with a different rule sequence.* ***A.*** *Description of the AABBAABBAA sequence.* ***B.*** *We adapted the rule learning index defined in the main text to measure the learning of the pairs. When the rule is not yet grasped by the network, the surprise vector shows a periodic pattern at the period of the pairs. We computed the FFT of the surprise signal and retrieved the power at the frequency of the pair. The evolution of the power at this specific frequency reflects how much the network is surprised at each switch.* ***C.*** *We found qualitatively similar results to what was found for the switch rule : the sparsity of the networks influenced the time needed for the rule discovery while the noisy recurrent connections influences bot the integration and the rule learning.*