# Reconstructing the Bayesian Adaptive Toolbox: Challenges of a dynamic environment and partial information acquisition

**Percy K. Mistry (pkmistry@uci.edu)**
**Jennifer S. Trueblood (jstruebl@uci.edu)**
Department of Cognitive Sciences, University of California Irvine, Irvine, CA 92697-5100 USA

**Abstract**

We show how dynamic (changing) environments can affect choice behavior, and highlight the challenges that recent models face in explaining the learning and selection of heuristic strategies under such conditions, especially when decisions are made using only a small subset of the available information. We propose an enhanced modeling framework that includes a trial-by-trial implementation of a Bayesian adaptive toolbox, redefinition of heuristic strategies, and incorporation of intricate learning rate mechanisms into a strategy learning model. We use data from a new empirical study to show how this improves the quality of inference.

**Keywords:** Learning; Bayesian graphical models; strategy selection; heuristics; adaptive toolbox; Bayesian inference; dynamic environments; reinforcement learning

## Introduction

We investigate the strategy selection problem in multiple cue probability learning (MCPL) tasks by applying Bayesian inferential approaches to the question of how strategy selection and learning can be investigated. Scheibehenne, Rieskamp, & Wagenmakers (2013) introduced a Bayesian adaptive toolbox where the probability of using each strategy is inferred across all trials rather than on a trial-by-trial basis.

We implement a similar toolbox approach on a trial-by-trial basis, and augment this with a hierarchical learning mechanism (strategy selection and learning (SSL), Rieskamp & Otto, 2006) that governs the shift in use of strategies over trials. This allows us to capture adaptive behavior under dynamic environmental conditions where people demonstrate substantial learning effects. Next, we show how the standard definition of heuristic strategies is ineffective in accounting for behavior in the unique class of experimental paradigms we have selected. To tackle this, we propose that such standard heuristic strategies be redefined at the level of their elementary building blocks (Gigerenzer and Gaissmaier, 2011) and incorporate Bayesian inference based belief updating[1] into the learning mechanism. Finally, we adapt the SSL model to incorporate more elaborate learning rate mechanisms such as variable learning rates, item-specific learning, random-effects, volatility dependent learning, and counterfactual learning.

We first describe a new empirical study that we use to subsequently show how our model can provide greater insight into adaptive behavior. Our study focuses on a forced choice paradigm where the environmental conditions can change over the course of trials, and where it is possible for participants to acquire only a partial subset of the cue information available.

## Experiment

We use a MCPL paradigm where participants had to make repeated forced choices between one of three options on the basis of a set of underlying cues. This paradigm was similar to that used by Bröder & Schiffer (2006) with changes in the conditions and relationships between cues and options.

### Methods

34 University of California, Irvine undergraduates participated in the experiment for course credit. The cover story for the choice task was a hypothetical stock market game, in which participants had to choose between three financial stock options, based on four possible binary cue attributes that included past profit growth, sales growth, and recommendations from two independent advisors. The cue values were instantiated as High / Low for the two financial indicators, and as Buy / Sell for the two advisory cues. Participants could acquire any of the twelve (four attributes x three options) cue values in any order, at a cost of 5% of the total gross reward obtained for that trial, for each cue acquired. Once a participant selected a cue attribute, it remained visible throughout the trial. Each participant made choices for 120 trials split into four blocks of 30 trials each. Each block was associated with either a compensatory (C-Block) or non-compensatory (N-Block) environment, which differed in the relationship between the cue values (c1 to c4, encoded as +1 or -1) and the associated reward outcomes (R), such that gross rewards ranged from –150 to +150:

$R(C) = 40c1 + 37c2 + 34c3 + 31c4 + noise(-8, 8)$

$R(N) = 78c1 + 7c2 - 21c3 - 36c4 + noise(-8, 8)$

The objective of the task was to maximize the rewards remaining after any cue acquisition related costs. In our task, a take-the-best (TTB) strategy always provided a higher net payoff in N-blocks and a weighted average (WADD) strategy in C-blocks. The actual cue weights, order of importance or validity of the weights, or the type of environment for each block was not known to the participants (this was to be learned as part of the decision process), but they were told that the underlying environment and relationships between cues and options could change between blocks, but remained constant within a block. The start and end of each block was clearly indicated. The cue weights included negative values (this design was implemented to nudge participants towards a more

[1] The learning mechanism is not changed to Bayesian learning, just the process by which the modeler determines what strategy is to be reinforced is mechanized as a Bayesian inference process

deliberative cognitive effort as opposed to focusing on the salience of the positive valued cues), and while the cue weights and order were not disclosed, participants were told that it was possible for cues to be negatively related to the options (brief examples of how this could be justified within the paradigm were included). After each trial, feedback was provided on the reward associated with all the three options.

The block size was designed to be small (30 trials each) to manipulate the possible effects of routinization of decision strategies (see Bröder & Schiffer, 2006; Bröder et al, 2013). The experiment consisted of a factorial design with four conditions (2 conditions starting either with a C-Block or N-Block x 2 conditions where the routine was manipulated by either alternating or placing similar blocks consecutively). Thus, the four blocks under the four conditions could be represented as CNCN, CCNN, NCNC and NNCC. We wanted to measure the interaction between routine length and starting conditions, manifested as the extent of maladaptive routinization, initial bias and response to different levels of volatility that the changing environment represented, assess the potential inadequacies of existing models to explain underlying behavior, and demonstrate how these models could be suitably improved.

## Results

Table 1 shows the result of Bayesian t-tests[2] for pairwise differences in the number of cues acquired, and the performance (standardized reward scores[3]), between conditions. This shows a hierarchy of performance levels, with significant evidence for higher performance in the CCNN condition, no significant difference between the two alternating conditions (CNCN and NCNC) as the t-test comparing these yields a BF in favor of the null, and reasonably strong evidence for a lower performance in the NNCC condition. The mean performance score reflects this trend (CCNN 0.86, CNCN & NCNC 0.78 and NNCC 0.73).

A Bayesian ANOVA test (Table 2) reveals that the routine type alone (consecutive vs alternating blocks) is not a significant factor, and the best model that explains the variance is based on the starting condition (C vs N) and a strong interaction between the starting condition and the routine type. Analyzing the coefficients for the factors under this model reveal an interesting relationship where the routine types have an asymmetric impact depending on the starting condition. Consecutive routines interacting with starting C have a coefficient of +0.03, and -0.03 when interacting with starting N; alternative routines are exactly the other way around. Starting conditions alone have a coefficient of +0.03 (C) and -0.03 (N), reflecting initial bias.

Table 1: Pairwise Bayesian t-test between conditions

| Difference | Performance | | Cue Acquisition | |
|---|---|---|---|---|
| | Log(BF) | Delta | Log(BF) | Delta |
| CNCN vs CCNN | **15.8** | -0.26 | **10.6** | -0.22 |
| CNCN vs NCNC | -3 | na | 2.8 | -0.15 |
| CNCN vs NNCC | 3.2 | 0.16 | **13.8** | -0.26 |
| CCNN vs NCNC | **15.3** | 0.27 | -2.2 | na |
| CCNN vs NNCC | **43.8** | 0.43 | -2.7 | na |
| NCNC vs NNCC | 2.2 | 0.15 | -1.1 | na |

Table 2: Bayesian ANOVA test for factors contributing to the standardized performance score

| Model vs Baseline (Intercept only) | Log(BF) |
|---|---|
| Routine (Consecutive vs Alternate) | -2.1 |
| Routine + Routine:Starting[4] | 15.1 |
| Routine + Starting | 16.2 |
| Routine:Starting[4] | 17.4 |
| Starting Condition (N vs C) | 18.2 |
| Routine + Starting + Routine:Starting[4] | 33.5 |
| **Starting + Routine:Starting[4]** | **36.1** |

We also test for differences within and between block types (see Table 3). There is strong evidence for within block improvement in performance between the first (HB1) and second (HB2) half of each block. This is true for both environmental conditions, but the effect size and significance is much higher in N-blocks vs C-blocks.

Table 3: Bayesian t-test (within & between blocks)

| Difference | Performance | | Cue Acquisition | |
|---|---|---|---|---|
| | Log(BF) | delta | Log(BF) | Delta |
| HB2 vs HB1 (C) | **39** | 0.15 | 3.2 | -0.11 |
| HB2 vs HB1 (N) | **193** | 0.32 | **74.6** | -0.41 |
| C vs N (overall) | -1.2 | na | **51.1** | 0.23 |
| C vs N (HB1) | 2.2 | 0.10 | 1.3 | 0.09 |
| C vs N (HB2) | -3.2 | na | **72.0** | 0.39 |
| C2 vs C1 (all) | **9.1** | 0.16 | 0.16 | -0.08 |
| N2 vs N1 (all) | **57** | 0.35 | **44.3** | -0.31 |

The overall performance of N-blocks and C-blocks however is not different (significance test yields a BF in favor of the null; overall accuracy for C is 82.3% and for N is 82.9% and standardized performance score for C is 0.80 and N is 0.78). Rather, a half-block comparison between C and N blocks shows that first half performance in C-blocks is marginally better than in N-blocks (log(BF)=2.2 in favor

[2] Results are summarized as the log(Bayes factor) in favor of a difference vs the null (zero difference), based on a JZS t-test, and the standardized mean difference (Delta) which shows the effect size. A log(BF) with absolute value < 1 can be considered inconclusive. Log(BF) > 1 indicates evidence in favor of the difference, and log(BF) < -1 in favor of the null. Larger log(BF) values imply greater evidence of a difference. Very large differences are highlighted in bold.

[3] We measured effective performance by normalizing rewards from 0 to 1 based on the maximum and minimum possible gross rewards on each trial.

[4] 'Routine:Starting' indicates the interaction effect between routine and starting conditions

of a difference). This leads to the conclusion that performance starts at a lower level in N-blocks, perhaps reflecting an initial bias, but seems adaptive enough to reach overall C-block levels. This adaptivity is also reflected in the cue acquisition levels, (average cue acquisition is C 33%, N 28%), where a t-test is not really conclusive for the first half of C and N blocks, but shows a very significant (log(BF) = 72) and strong effect size for higher cue acquisition levels in the second half for C-blocks vs N-blocks, a significant (log(BF) = 74.6) and strong reduction in cue acquisition between the first and second half within N-blocks, and a significant increase in performance between the first and second encountered blocks of the same type, with a stronger effect for N-blocks. Our modeling efforts attempt to capture this behavior via inference on the underlying latent heuristic strategies utilized by participants, how these strategies change with changing environmental types, and the differences between conditions.

## Modeling Challenges and Enhancements

### What to reinforce? A Bayesian solution

Most learning algorithms update beliefs about a set of items under consideration. If the locus of learning is a choice option, learning can be explicitly modeled since the selected choice option is always known. When the locus of learning is a latent process (in our case, a heuristic or strategy that cannot be directly observed), the modeling process needs to infer which latent item was utilized and hence needs to be reinforced or updated by the learning algorithm on each trial. Existing approaches to identify such latent strategies include response matching (strategy prediction matches the actual response observed; see Rieskamp, 2008) or an additional criterion of minimum cue acquisition (minimum cues required to implement the strategy have to be acquired; see Rieskamp & Otto, 2006).

In paradigms, like ours, which allow partial information utilization and where the cue acquisition density is very low, response matching alone is unrealistic, since the number of cues acquired are rarely adequate (e.g. in our study, average cue acquisition levels were 31%, with over 50% of cues being acquired on only about 12% of the trials) to implement standard heuristic strategies (e.g. TTB, WADD, tallying, and so on). Including a criterion for minimum cue acquisition makes most of the updates ineffective, since none of the strategies would be updated on a large number of trials (e.g. in our study, updates on 90% of the trials become ineffective since they do not meet the cue acquisition criteria for any commonly defined strategy). To counter these issues we propose a possible solution, partially redefining what is considered a 'strategy', as part of an adaptive toolbox of strategies. The existing SSL model calculates the probability of using each strategy ($s_i$) on each trial (t) based on the underlying value assigned to each strategy, called q-values ($p(s_i,t) = Q(s_i,t) / \Sigma_j Q(s_j,t)$). The q-values are updated based on the observed choice ($c_o$) and associated reward ($Q(s_i,t) = Q(s_i,t-1) + I(s_i,t-1)*r(c_o,t-1)$), where $I(s_i,t)$ is an indicator function based on response matching or response matching and minimum cue acquisition, that indicates whether a strategy was used on a particular trial. The initial q-values depend on an initial association parameter (K), initial strategy preference ($\beta$) and the maximum possible reward on any trial ($Q(s_i,1) = \beta_i * K * R_{max}$). We propose modifying the q-value calculation for each strategy to $Q(s_i,t) = Q(s_i,t-1) + p(s_i,t|c_o,u) * r(c,t-1)$, where we define $p(s_i,t|c_o,u)$ as the Bayesian posterior probability of the participant having utilized a specific strategy ($s_i$) on trial (t), based on the observed choice ($c_o$) and the pattern of cues acquired (u):

$$p(s_i,t|c_o,u) = \frac{p(c_o,t|u,s_i)\, p(u,t|s_i)\, p(s_i,t)}{p(c_o,t|u)\, p(u,t)}$$

Here, $p(s_i,t)$ is the prior probability of utilizing a strategy ($s_i$) on trial (t) as predicted by the cognitive model. The remaining probabilities, $p(u,t|s_i)$, the probability of acquiring the specific cue pattern given the application of a specific strategy, and $p(c_o,t|u,s_i)$, the probability of making a specific choice given the particular strategy being used and the cue pattern acquired, need to be specified. We propose that these probabilities be defined as the information search and decision rule building blocks of the heuristic strategy.

### Redefining heuristic strategies

Similar to traditionally defined strategies (TS), $p(c_o|u,s)$ is simply a decision rule that combines all the cue information in exactly the same way, but taking into account only the cue values that have actually been acquired on each trial. If the decision rule applied to the partially observed cues can discriminate between all options, this probability is either 0 or 1 for each option (c), otherwise it is distributed across the non-discriminable options. We call these strategies observed-cue strategies (OS), to differentiate the level at which the decision rule is applied. On the other hand, defining p(u|s) at the level of a heuristic strategy is quite different from the traditional cue search rules. We propose that the cognitive act of cue acquisition can be envisaged as a sampling process, and the heuristic defines the probability distribution of cue acquisition patterns from which the individual is sampling. To implement this, we categorize each possible pattern into one of a number of ordered subset categories on the basis of identified classifiers. The extreme categories define a pair of complementary approaches to cue acquisition, and the ordered categories are defined to represent a log-odds ratio between the two complementary approaches (i.e. extreme categories include cue acquisition patterns that have extreme log-odds ratios and those in the middle reflect patterns that are equally likely under the two acquisition approaches). The log-odds for each ordered category are derived from the classifiers by calculating an index score for each category. For our study, we categorized all possible (4096) cue acquisition patterns by using a simple classification scheme with three classifiers: (1) the number of unique cue attributes where at least one cue value was selected (higher value indicates a compensatory approach), (2) the cue acquisition density within this subset

of attributes (lower indicated a greater propensity to compare cues across attributes, thus compatible with a compensatory approach), and (3) an assumption of sensitivity to costs (this redistributed the probabilities for each approach towards patterns with lower cue acquisition density). Using these classifiers we could obtain a formulaic characterization of different cue acquisition patterns yielding a score of 0 to 1 for each[5], which was then scaled to reflect the log-odds, and transformed to a probability using the inverse logit function. We could thus define a pair of heuristic approaches (approximated as compensatory and non-compensatory) with complementary probability distributions over all possible cue patterns, defining two sets of 'p(u|s)'. While further details of this mechanism are beyond the scope of this paper, we highlight that the classifiers used were not exhaustive, but an illustrative instantiation of a working model for our toolbox.

## Modeling Learning Rates

Previous implementations of SSL usually assume a constant learning rate across all trials, often parameterized as an initial association level (but see Gluth, Rieskamp, & Buchel, 2014 for a fixed learning rate implementation). However, learning rates might be influence by a number of different factors. Thus, we explore four different learning rate mechanisms: (1) variable, (2) entropy-based learning rates, (3) random effects, and (4) counterfactual learning.

**Variable Learning Rates (SSL-V):** Studies have shown that flexible and variable learning rates within RL based mechanisms can improve predictions and also produce results comparable to Bayesian learning (Payzan-LeNestoury & Bossaertsz, 2014; Speekenbrink, & Konstantinidis, 2014). We re-parameterize the SSL model to include a flexible learning rate that can depend on the environmental condition (i.e. change between experimental blocks) and can also be strategy-dependent. The latter formulation can be interpreted as responses to different levels of 'surprise' to the same reward outcome that different strategies generate. The learning rate for each block type (b) and strategy ($s_i$), $L(s_i,b)$, is modeled using a hierarchical prior, and the initial association parameter (K) is no longer required. The revised q-value calculation is $Q(s_i,t) = Q(s_i,t\text{-}1) + p(s_i,t|c_o,u) * r(c_o,t\text{-}1) * L(s_i,b)$.

**Entropy-based Learning Rates (SSL-E):** Allowing the learning rate to vary between blocks and strategies still enforces a fixed learning rate within blocks. Assessment of environmental volatility and detection of environmental changes have been implicated in the modulation of learning rates (Pearson & Platt 2013; Behrens et al, 2007). We propose that the conflict between probabilities of using different strategies generated by the cognitive model can be interpreted as a possible proxy measure of the volatility. We implement a version of the model that modulates the learning rate on a trial-by-trial basis based on a moving average of recent entropy. Higher entropy reflects greater uncertainty in the environment and hence increases learning rates. Entropy (for 'n' strategies in a toolbox) is calculated based on the strategy probabilities generated by the model:

$$\text{Entropy}(t,n) = -(1/\log_e n) * \Sigma_{i=1:n}\{p(s_i,t) * \log_e(p(s_i,t)\}$$

**Random effects (SSL-R):** Since learning rate may be subject to individual (participant) effects, item effects (individual strategies, type of environment – compensatory or non-compensatory) as well as the experimental conditions, we propose a model where these 4 main effects are broken down as random effects and combine additively on a probit scale (see Rouder & Lu, 2005 for Bayesian modeling of crossed random effects). We develop a version of our model assuming independent random effects, and each of these (L_ind, L_strat, L_env, L_cond) are modeled hierarchically with a scaled normal prior and a hyperparameter for the effect standard deviation. The posterior distribution of the standard deviation indicates the level of heterogeneity arising from each effect.

$$L\_probit = L\_mean + L\_ind + L\_strat + L\_env + L\_cond$$
$$L_r(s_i,b) = \text{Scaling Factor} * \Phi\,(L\_probit)$$

**Counterfactual Learning (SSL-C):** Counterfactual learning of choice options is commonly studied, however incorporating it into a task paradigm with low cue acquisition density and a latent locus of learning can get tricky, as counterfactual implications of traditionally defined compensatory strategies when actual behavior is non-compensatory cannot be easily evaluated. We can however implement counterfactual learning successfully using our approach to defining heuristics based on observed cues, as below, where $I_{ij}$ is an indicator function that reflects whether a particular strategy predicts choice '$c_j$', given the observed cues, and CF is a free parameter [0, 1] indicating the relative strength of counterfactual learning.

$$Q(s_i,t) = Q(s_i,t\text{-}1) + \Sigma_j \{ I_{ij} * r(c_j,t\text{-}1) * L(s_i) * [p(s_i,t|c_j,u)+(1\text{-}p(s_i,t|c_j,u))* CF] \}$$

## Bayesian inference framework

The learning model is implemented in a Bayesian inference framework that allows hierarchical estimation of the parameters. The calculated strategy probabilities are converted to choice probabilities for each choice option 'j', and include an application error rate (AER), $\epsilon_i$, defined independently for each strategy 'i'.

$$p(c_j) = \Sigma_i \{ p(c_j|s_i)(1\text{-}\epsilon_i) + (1\text{-}p(c_j|s_i))\,\epsilon_i / (N\text{-}1) \}$$

Here, N is the number of different choice options. The AER captures variability in the decision rule but cannot account for variability in the cue acquisition process. To ensure that the sum of posterior probabilities inferred in the

[5] For instance, selecting all 3 cue values corresponding to the three options from a single attribute, representing a non-compensatory approach, resulted in a raw score of 0.08, whereas selecting 3 cue values, each from a unique attribute-option pair, thus representing a compensatory approach, resulted in a raw score of 0.75.

Bayesian inference for belief updating sum to one over all strategies, we propose the inclusion of a guessing strategy, which defines an equal probability distribution over all choices and cue acquisition patterns, and is reinforced with a probability $1 - \Sigma_i\ p(s_i,t|c_o,u)$. This allows the model to capture variability in cue acquisition behavior that cannot be reasonably explained by any of the strategies.

## Modeling Results

Table 4 summarizes the comparative performance of models including a static toolbox and an unchanged SSL model based on traditional strategies (TS), response matching and minimum cue acquisition (CA), and our proposed model based on observed strategies (OS), Bayesian updating (BU) and four models with alternate learning rate structures (SSL-V, SSL-E, SSL-R, SSL-C). All models were built including a compensatory (CS), non-compensatory (NS) and guessing strategy. We use the human data from the described experiment to generate a posterior distribution of the parameter values and a posterior predictive distribution of the data. The posterior predictive distribution reveals the distribution over all possible data points that the model predicts based on the inferred posterior distribution of the parameters after having seen the data. All of the following analysis is thus based on Bayesian inference on the observed human data using the above described cognitive models. Our framework (OS / BU / modified learning rates) provide higher accuracy (Acc) of the posterior predictive (thus, the best account of the observed human data), improved (lower) deviance information criteria (DIC), and better qualitative insights compared to existing approaches (Static / TS / CA) as we demonstrate in the subsequent commentary.

Table 4: Model Comparison

| Learning Rate | Strategy type / Update | | Acc | DIC | Insight b/w conds | Insight b/w blocks |
|---|---|---|---|---|---|---|
| Static | TS | - | 72% | 4379 | Limited | No |
| SSL | TS | CA | 72% | 4413 | Limited | No |
| Static | OS | - | 77% | 3277 | Limited | No |
| SSL-V | OS | BU | 80% | 2786 | Yes | Yes |
| SSL-E | OS | BU | 80% | 2950 | Yes | Yes |
| SSL-R[6] | OS | BU | 80% | 2915 | Yes | Yes |
| SSL-C | OS | BU | 82% | 2745 | Yes | Yes |

The static (no learning) toolbox model using traditional strategies (TS) predicted an 85% use of CS and practically no usage of NS. Incorporating an SSL mechanism using response matching and minimum cue acquisition (CA) worsened the DIC even further, primarily because cue acquisition was unable to account for any of the standard strategies on most trials, and 90% of the updates were ineffective. This is reflected in the lack of coherence between the high probability (> 80%) of CS predicted by this model and CS being updated by the learning model for only 0.5% of the trials. Differences between conditions are explained via minor differences in the probability of guessing, with no insight into differences between blocks.

Next, we assessed the static toolbox without learning using the observed-cue decision rules (OS) and found that this outperformed the previous models considered in terms of accuracy and DIC. This also provided more realistic application error (AER) rates (NS dropped from 28% to 10%, CS fell from 9% to 0.5%), and a more balanced view of the average strategy usage (55% CS, 35% NS). It also infers that condition NNCC has the lowest usage of CS (possibly implying some form of routinization) and the highest guessing rate (22%), with the CCNN and NCNC both having similar high rates of CS and lowest rate of guessing (2-3%), and CNCN lying in between these.

Implementing our revised learning model (SSL-V) provided inference on average strategy usage similar to the static mode, but improved accuracy and DIC even further. Inferences from the entropy (SSL-E), counterfactual (SSL-C), and measurement of individual differences and random effects (SSL-R) models provided similar estimates, although only SSL-C model improved fit compared to the SSL-V model. In all of these models, the ineffective updates reduced from 90% to only about 15%, which contributes significantly to the improved performance of the models. The updates for individual (NS and CS) strategies are now also coherent, being in the range of the strategy use predicted by the model (unlike the CA implementation).

Most importantly, these models now provided a dynamic account of how strategy use shifted on a trial-to-trial basis, within and between blocks, and between conditions. All the OS-based SSL adaptations provide a similar insight into the dynamics of strategy use across blocks. Figure 1 shows the strategy usage inferred on a trial-by-trial level from one of these models. In the CNCN condition, usage of CS strategy is well-tuned to the advent of C-blocks, but there is a considerable amount of guessing in the N-blocks. While the participants seem to be picking up on the differences in the alternating blocks, finding the right NS seems to be harder.

In the CCNN condition, participants seem to recognize the change in the first N block which is quite volatile in terms of use of strategies, but interestingly, they revert back to the previously routinized CS strategy within this block itself, and implement it even more strongly in the last N-block. Once again, this shows change detection at the advent of the N-block, and a similar difficulty in finding the right NS. But it seems that the higher routinization of CS makes participants revert back to CS rather than adopt a guessing strategy, which is rarely used in this condition, hence leading to the best overall performance. In the contrasting NNCC condition, participants again start with an initial preference for CS, but this not seem to be strongly

[6] Bayesian inference was carried out using MCMC sampling. MCMC chain convergence was good ($R<1.1$) across parameters for all models considered, except SSL-R, where a few individual parameters showed poor convergence ($R>1.1$). We restrict analysis primarily to the models where convergence was not an issue.

reinforced and is replaced with a higher usage of a guessing strategy, until the advent of the first C-block. In the NCNC condition however, participants do not revert to guessing in the first block and their use of a CS strategy is positively reinforced, resulting a stable preference across blocks. Comparing performance in N1 between NNCC and NCNC conditions reveals that the better reinforcement of CS in NCNC is a result of the cue patterns selected, resulting in the best option being selected 80% of the time, vs 65% in NNCC. Relating this to our redefined strategies, future analysis could consider comparing these conditions with different distributions from which cue patterns are sampled.

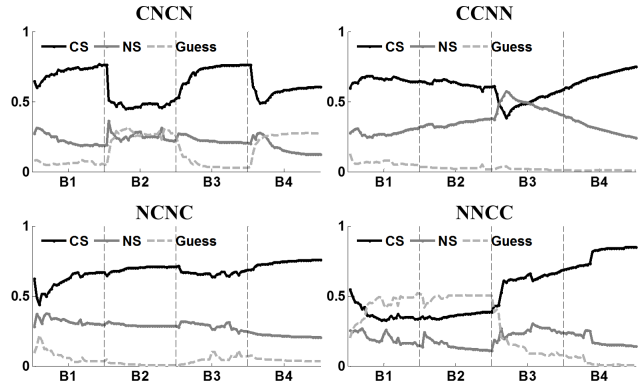


Figure 1: SSL-V model: Mean probabilities of strategy use across trials (B1 to B4 are four blocks of 30 trials each)

Also interesting were the learning rates inferred by these models. The SSL in its original form was modeled using the initial association parameter, which can be interpreted as the inverse of the learning rate (but not exactly equivalent). Implementing SSL using only response matching (RM) inferred a low initial association of 20 (previous studies have yielded best fit parameters in the range of 50-100), whereas a more realistic implementation incorporating minimum cue acquisition as well inferred an extremely high value of 1047, thus inferring almost no learning (since 90% of the trials were ineffective learning updates). Implementing our revised models using a re-parameterized learning rate yielded an average learning rate of 1.3 (SSL-V), 0.4 (SSL-C), 0.5 (SSL-E), and 2.1 (SSL-R). The counterfactual model (SSL-C) includes a larger breadth of learning, and the entropy model (SSL-E) typically predicts frontloading of the learning rate, which gradually drops and settles to lower levels as entropy is resolved. SSL-C also infers that the extent of counterfactual learning (inferred parameter CF) is lower in the NNCC (0.36) condition as compared to the remaining conditions (average 0.49).

Interestingly, while the SSL-V model shows a higher learning rate for CS as compared to NS, segregating these effects as random effects in the SSL-R model reveals a more intricate pattern. It reveals that N-blocks and NS strategies contribute to higher learning rate effects than C-blocks and CS strategies respectively, and also that maximal heterogeneity is observed in individual participant and strategy type effects. This ties in nicely with the empirical observations on differences in adaptivity within N-blocks being higher than for C-blocks. It also explains the higher volatility of probabilities of strategies within N-blocks.

## Conclusion

We implemented an experimental design to identify behavioral patterns in a paradigm where the environmental conditions change and the information costs push participants towards partial information acquisition. We demonstrated how otherwise successful models may be rendered inadequate, and successfully built a computational framework to reconstruct a Bayesian adaptive toolbox, improving our ability to account for observed behavior.

## References


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.

Bröder, A., Glöckner, A., Betsch, T., Link, D., & Ettlin, F. (2013). Do people learn option or strategy routines in multi-attribute decisions? The answer depends on subtle factors. *Acta psychologica*, 143(2), 200-209.

Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 32(4), 904–918.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62, 451-482.

Gluth, S., Rieskamp, J., & Buchel, C. (2014). Neural Evidence for Adaptive Strategy Selection in Value-Based Decision-Making. *Cerebral Cortex*, 24(8), 2009–2021.

Payzan-LeNestour, E., & Bossaerts, P. (2014). Learning about unstable, publicly unobservable payoffs. *Review of Financial Studies*, hhu069.

Pearson, J. M., & Platt, M. L. (2013). Change detection, multiple controllers, and dynamic environments: Insights from the brain. *Journal of the Experimental Analysis of Behavior*, 99(1), 74–84.

Rieskamp, J., & Otto, P. E. (2006). SSL: a theory of how people learn to select strategies. *Journal of Experimental Psychology: General,* 135(2), 207.

Rieskamp, J. (2008). The importance of learning when making inferences. *Judgment and Decision Making*, 3(3), 261–277.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12(4), 573-604.

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E. J. (2013). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review*, 120(1), 39.

Speekenbrink, M., & Konstantinidis, E. (2014). Uncertainty and exploration in a restless bandit task. *Cognitive Science Society*.