What is the nature of decision noise in random exploration?

Siyu Wang[1] and Robert C. Wilson[1,2]

[1]Department of Psychology, University of Arizona, Tucson AZ USA

[2]Cognitive Science Program, University of Arizona, Tucson AZ USA

December 4, 2017

Abstract

The explore-exploit tradeoff is a fundamental behavioral dilemma faced by all adaptive organisms, from everyday life decisions like deciding for a meal to important life decisions like finding a life partner. It is computationally a hard problem to find the right balance between exploration and exploitation and hence there is significant interest in how humans and other animals solve the explore-exploit dilemma. One particularly effective strategy for solving the explore-exploit dilemma is choice randomization. In this strategy, the decision process is noisy meaning that high value 'exploit' options are not always chosen and exploratory choices are sometimes made by chance. In theory, such 'random exploration', can be surprisingly effective in explore-exploit problems and, if implemented correctly, can come close to optimal performance. Recent work suggests that humans actually use random exploration to solve simple explore-exploit problems. Despite this progress a number of questions remain about the nature of random exploration as there are a number of ways in which seemingly stochastic choices could be generated. In one strategy, that we call the external noise strategy, participants could rely on stochasticity in the world and allow irrelevant features of the stimulus to drive choice. In another strategy called internal noise strategy, people could rely on stochastic processes within their own brains. In this work, we modified our recently published 'Horizon Task' in such a way as to distinguish these two strategies. Using both a model-free and model-based analysis of human behavior we show that, while both types of noise are present in explore-exploit decisions, random exploration is dominated by internal noise. This suggests that random exploration depends on adaptive noise processes in the brain which are subject to (perhaps unconscious) cognitive control.

Introduction

Imagine trying to decide where to go to dinner with a friend, you can go to your favorite restaurant that you both really enjoy and always go to, or you can try the new restaurant that just opened a few days ago right across the street to the other restaurant which may end up being your newly favorite. More generally, such decisions are known as explore-exploit decisions. The explore-exploit decision refers to deciding between exploiting the best known option so far, like going to your old favorite restaurant, and exploring other options for potential better decisions in the future, like trying the new restaurant. There's considerable interest in how humans and animals solve it. (CITE 10 PAPERS INCLUDING WILSON ET AL. 2014...)

One particularly effective strategy for solving the explore-exploit dilemma is choice randomization (). In this strategy, the decision process between exploration and exploitation is corrupted by 'decision noise', meaning that high value 'exploit' options are not always chosen and exploratory choices are sometimes made by chance. In our restaurant example, your restaurant decision does not always depend on the quality of the food, you are very likely to go to the new restaurant if you happen to see another old friend walking right in, or you wait until the last moment to make a split second decision about where to go as if you flipped a mental coin in your head and decide to go to your old favorite if it's heads up. In theory, such random exploration, is surprisingly effective and, if implemented correctly, can come close to optimal performance ().

Recently we have shown that humans appear to actually use random exploration and actively adapt their decision noise to solve simple explore-exploit problems (). The key manipulation in the task is the horizon condition, i.e. the number of decisions remaining to make. The idea behind this manipulation is that people should explore more in the long horizon condition. If you are leaving town tomorrow for vacation, you probably want to go to your old favorite to guarantee a good last meal, but if you are not going anywhere, you would be more likely to try the new restaurant. Using such a horizon manipulation we found that people have greater decision noise in the long versus the short horizon condition.

However, a key limitation of this work was that the source of the decision noise used for exploration is unknown. Of particular interest is whether the adaptive decision noise that is linked to exploration is generated internally, within the brain, or arises externally, in the input from the world. In the restaurant example, an old friend walking by would be a source of external noise, but flipping a mental coin would be an internal noise. Previous work makes a strong case for both types of noise being relevant to behavior. For instance, external, stimulus-driven noise is thought to be a much greater source of choice variability in

perceptual decisions than internal noise (). Conversely internal, neural noise is thought to drive exploratory singing behavior in song birds () and the generation and control of this internal noise has been linked to specific neural structures.

In this paper, we investigate which source of noise, internal vs external, drives random exploration in humans in a simple explore-exploit task adapted from our previous work (). To distinguish between the two types of noise, we had people make the exact same explore-exploit decision twice. If decision noise is purely externally driven, then people choices should be identical both times, that is their choices should be consistent since the stimulus is the same both times. Meanwhile, if noise is internally driven, the extent to which their choices are consistent should be determined by the level of the internal noise. By analyzing behavior on this task in both a model-free and model-based manner, we were able to show that, while both types of noise are present in explore-exploit decisions, the contribution of internal noise to random exploration far exceeds that contributed by the stimulus.

Results

The Repeated-Games Horizon Task

We used a modified version of our previously published 'Horizon Task' (Figure **??**) to show the influence of internal vs external noise on people's decisions (). The key manipulation was to use repeated games to let people make the same decision twice. In the restaurant example, if your decision is mainly driven by external noise, then a few months later when you see the friend walking in front of you into the restaurant, you are very likely to make the same decision and follow him into that restaurant and says hi. However, if your decision is mainly driven by internal noise, then next time you make a split-second decision about where to go, you are equally likely to go to the other restaurant.

More specifically, we look at the contribution of external and internal noise by providing with participants the same decision problem twice in the task. In this task, participants play a set of games in which they make choices between two slot machines (one-armed bandits) that pay out rewards from different Gaussian distributions. To maximize their rewards in each game, participants need to exploit the slot machine with the highest mean, but they cannot identify this best option without exploring both options first.

In each game they made multiple decisions between two options. Each option paid out a random reward

between 1 and 100 points sampled from a Gaussian distribution. The means of the underlying Gaussian were different for the two bandit options, remained the same within a game, but changed with each new game. One of the bandit will always have a higher mean than the other. Participants were instructed to maximize the points earned over the entire task.

The first four trials of each game were forced-choice trials, in which only one of the options was available for the participant to choose. We used these forced-choice trials to manipulate the relative ambiguity of the two options, by providing the participant with different amounts of information about each bandit before their first free choice. The four forced-choice trials set up two uncertainty conditions: unequal uncertainty(or [1 3]) in which one option was forced to be played once and the other three times, and equal uncertainty(or [2 2]) in which each option was forced to be played twice. After the forced-choice trials, participants made either 1 or 6 free choices (two horizon conditions). In unequal uncertainty condition, people are more likely to choose the option that they know less about - the more informative option - to explicitly explore that option more. This type of information driven exploration is known as directed exploration.

These conditions allow us to measure directed and random exploration in a model-free way. Directed exploration is measured as the probability of choosing the more informative option in [1 3] condition whereas random exploration is measured as the probability of choosing the low mean option in [2 2] condition. In line with previous results, we showed that both directed and random exploration increase with horizon. Directed exploration is considered to be driven by information bias and random exploration is considered to be driven by decision noise, in this work, we are investigating which source of decision noise, external vs internal, drives random exploration.

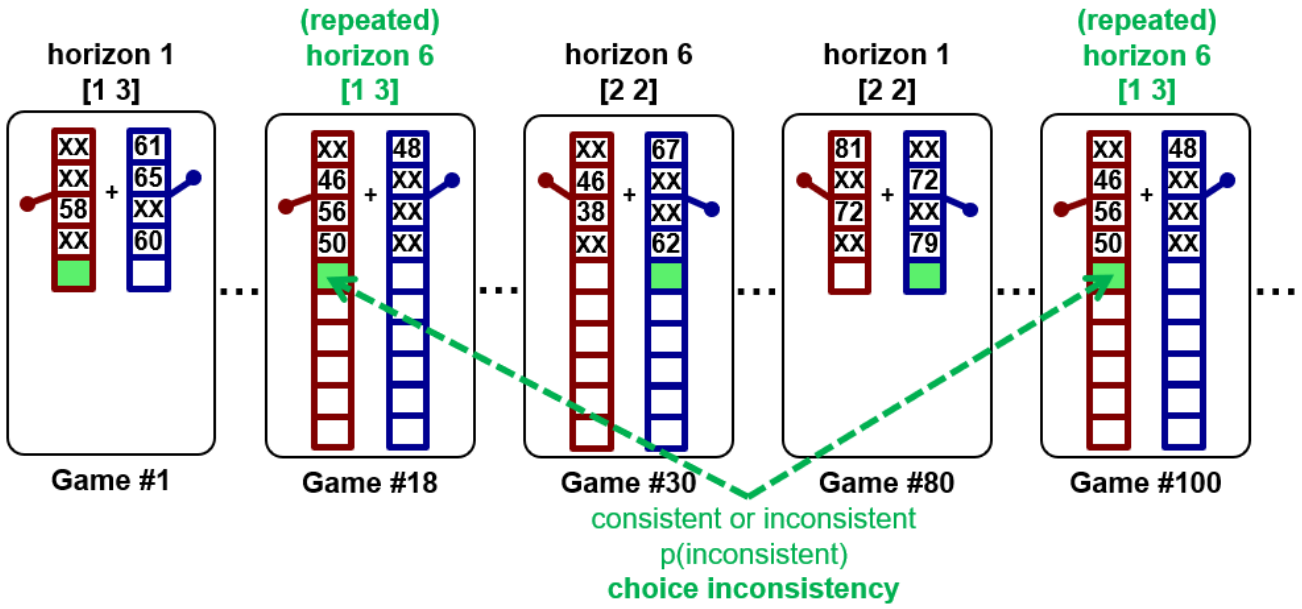


Figure 1: A key additional manipulation here is repeated games. Each pair of repeated games with identical example trials will appear twice during the experiment. We setup the repeated games such that they are at least 5 games apart from each other. A model-free measure of choice inconsistency which reflex the underlying decision noise is defined as the proportion of inconsistent choices for repeated games.

Finally, the crucial additional manipulation in this task is repeated games (Figure 1). In each pair of repeated games, the four forced-choice trials were yoked, meaning that on the first free choice trial participants were faced with identical stimuli. After the first free choice trial, the outcomes on the repeated games were not yoked and the outcomes were sampled independently from the underlying Gaussian distribution. Not yoking the later trials made it harder for participants to detect repeated games. In addition, the presentation of repeated games was controlled so that each repeated pair was at least five games away from each other.

Random exploration is dominated by internal noise

In this section we use both model-free and model-based analyses to show that both internal and external noise contribute to the behavioral variability in random exploration. Using the model-based hierarchical Bayesian analysis, we also show that the effect of internal noise is the main source of noise in random exploration.

Model-free analysis

In the model-free analysis we asked whether participants' choices were consistent or inconsistent in the two repetitions of each game. The idea behind this measure is that purely external noise should lead to consistent choices as the external stimulus is identical both time. Conversely, internal noise should lead to independent choices, and hence possible inconsistent choices both times. More specifically, we look at the proportion of times that a participant make inconsistent decisions in repeated game.

In addition, whether participants make consistent choices in repeated games is used as a model-free measure of internal noise, since in repeated trials, external noise should be identical on both trials. So only internal noise can differ and drive the choice inconsistency. The degree to which people make consistent choices in repeated trials can reflect the internal noise. Since choice inconsistency increases with horizon, this is a behavioral evidence that internal noise increases with horizon.

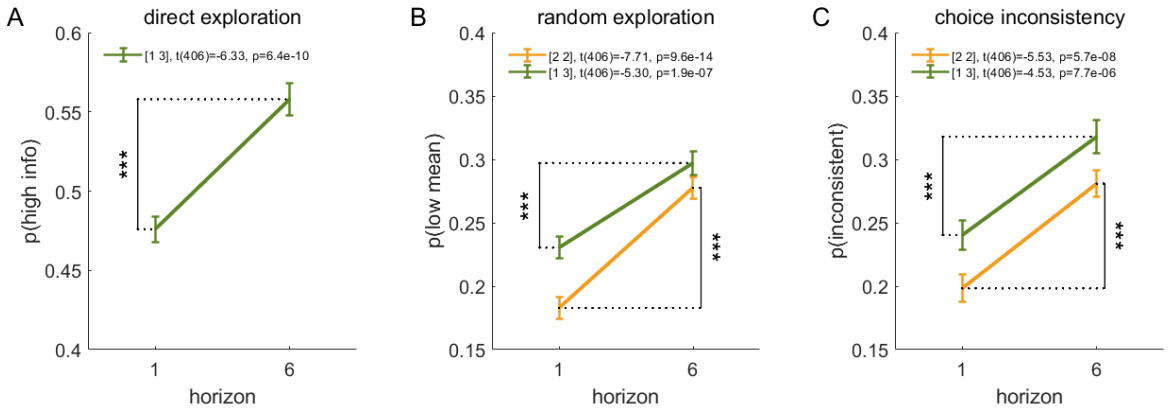


Figure 2: Both direct (t(406)=-6.33, $p < 0.001$) and random exploration (t(406) = -5.30, $p < 0.001$ for [1 3], t(406) = -7.71, $p < 0.001$ for [2 2]) increases with horizon.(A,B) Choice consistency in both [1 3] and [2 2] condition increases with horizon($P([1\ 3]) < 0.001$, $P([2\ 2])<0.001$).(C)

Choice inconsistency between repeated games was non-zero in both horizon conditions, suggesting that not all of the noise was stimulus driven. In addition, choice inconsistency was higher in horizon 6 than in horizon 1 for both [1 3] and [2 2] condition (Figure 2), suggesting that at least some of the horizon dependent noise is internal.

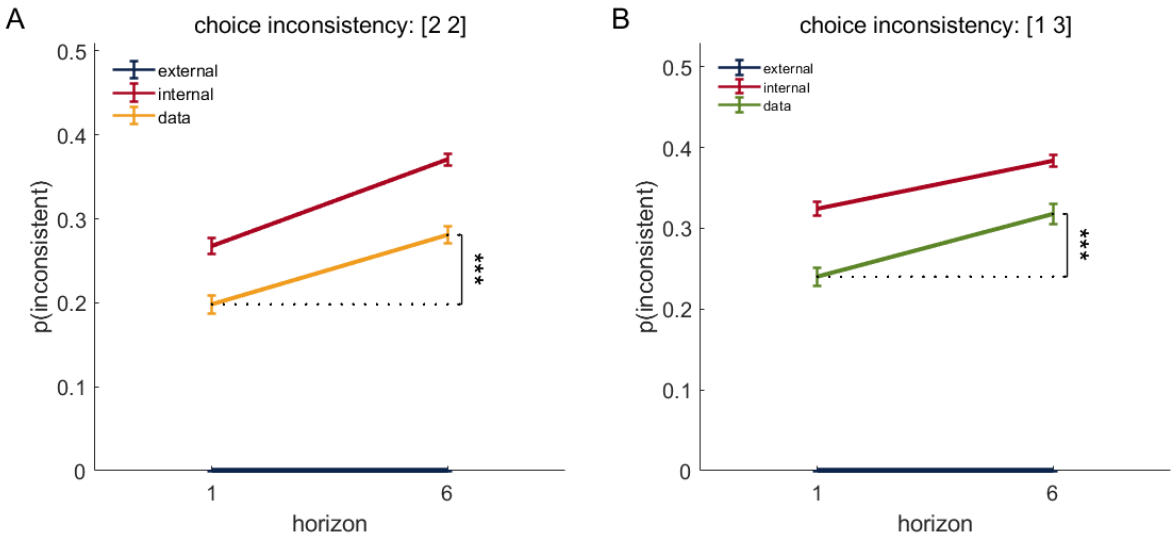


Figure 3: For both [1 3] and [2 2] condition, there is a significant difference between people's behavior and predicted choice inconsistency assuming that only external noise exists where people should behave identically in repeated games. Also, there is a significant difference between people's behavior and predicted choice inconsistency assuming that only internal noise exists where people treat repeated games independently.

To gain more quantitative insight into the data we computed predicted values of the choice inconsistency for the purely external and purely internal noise cases. For the purely externally driven noise case, then people should make the exact same decisions each time in repeated games, which corresponds to the light blue line in Figure 3. If noise is however purely internally driven, then there should be no stimulus-dependent noise component such that people should in principle treat repeated games independently, hence

$$\begin{aligned} P(\text{consistency}) &= P(\text{low mean})^2 + P(\text{high mean})^2 \\ &= P(\text{low mean})^2 + (1 - P(\text{low mean}))^2 \\ \text{hence,} \quad P(\text{inconsistency}) &= 1 - P(\text{consistency}) \end{aligned}$$

,

we can make the prediction of p(disagree) if there is only internal noise. However, people's behavior falls in between the pure external noise prediction and the pure internal noise prediction, suggesting that both external and internal noise are present in driving this choice inconsistency.

Since choice inconsistency only reflects internal noise, Figure 3 suggests that internal noise increases with horizon, however we can not draw any obvious conclusions about whether external noise is horizon

dependent or not.

Model-based analysis

To more precisely quantify the size of internal and external noise in this task, we turned to model fitting.

Overview of model As with our model-free analysis, the model-based analysis focuses only on the first free-choice trial since that is the only free choice when we have control over the information bias between the two bandits. To model participants choices on this first free-choice trial, we assume that they make decisions by computing the difference of value ΔQ between the right and left options, choosing right when $\Delta Q > 0$ and left otherwise. Specifically, we write

$$\Delta Q = \Delta R + A\Delta I + b + n_{ext} + n_{int} \tag{1}$$

Where, the experimentally controlled variables are $\Delta R = R_{right} - R_{left}$, the difference between the mean of rewards shown on the forced trials, and ΔI, the difference information available for playing the two options on the first free-choice trial. For simplicity, and because information is manipulated categorically in the Horizon Task, we define ΔI to be +1, -1 or 0: +1 if one reward is drawn from the right option and three are drawn from the left in the [1 3] condition, -1 if one from the left and three from the right, and in [2 2] condition, ΔI is 0). n_{ext} and n_{int} are external noise and internal noise respectively.

The subject-and-condition-specific parameters are: the spatial bias, b, which determines the extent to which participants prefer the option on the right, the information bonus A, which controls the level of directed exploration, n_{ext} denotes the external, external noise, which is identical on the repeat versions of each game, and n_{int} denotes internal noise, which is uncorrelated between repeat plays.

For each pair of repeated games, the set of forced-choice trials are exactly the same, so the external noise, n_{ext}, should be the same while the internal noise, n_{int} may be different. This is exactly how we distinguish external noise from internal noise. In symbolic terms, for repeated games i and j, $n^i_{ext} = n^j_{ext}$ and $n^i_{int} \neq n^j_{int}$.

Model fitting We used hierarchical Bayesian analysis to fit the parameters of the model (see Figure 7 for an graphical representation of the model in the style of CITE). In particular, we fit values of the information bonus A, spatial bias B, variance of internal noise σ^2_{int}, and variance of external noise, σ^2_{ext} for each participant in each horizon and uncertainty condition. The mean and standard deviation of information

bonus A and spatial bias B are sampled from a Gaussian prior and an exponential prior respectively. The variance for both type of noises were sampled from a gamma distribution, and the group-level parameter k and λ for the gamma distribution are sampled from exponential priors.

To capture the idea that external noise should be identical on repeated games, we sampled one value of the external noise, n_{ext} for each pair of repeated games. Conversely, because internal noise is expected to change between games we sampled two values of the internal noise, n^1_{int} and n^2_{int}, i.e. one for each individual game.

The model in Figure 7 is fitted using the MATJAGS and JAGS software ().

Parameter	Horizon dependent?	Uncertainty dependent?	Game dependent?
information bonus, A	yes	n/a	no
spatial bias, B	yes	yes	no
external decision noise, σ_{ext}	yes	yes	no
internal decision noise, σ_{int}	yes	yes	yes

Table 1: Model parameters.

Model fitting results Group-level estimates for the variances of internal and external noise are shown in Figure 6A. While both variances are non-zero, this shows that the internal noise is much larger than external noise. This horizon-based change is probed further in Figure XXXB in which we plot the posterior distributions over the change in internal and external noise with horizon. This clearly shows that only internal noise varies with horizon (STATS XXX% of the samples above zero for internal noise, XXX% of samples above zero for external noise).

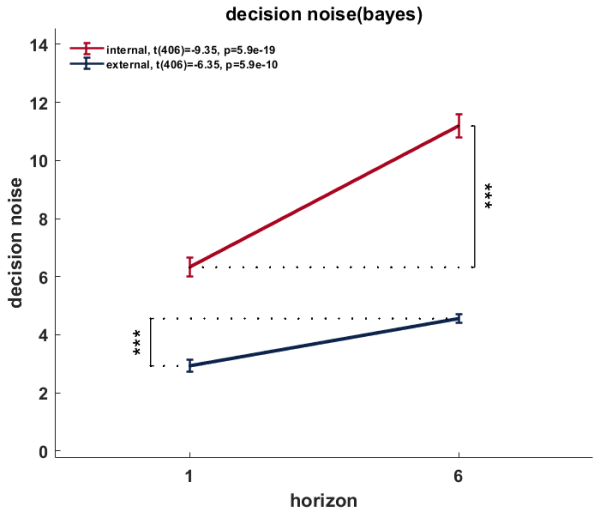


Figure 4: Both internal and external noises increase significantly with horizon.

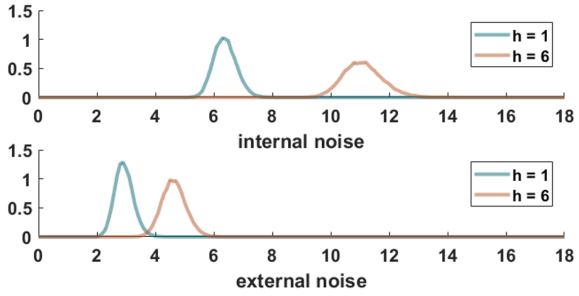


Figure 5: The posterior distribution of internal and external decision noise.

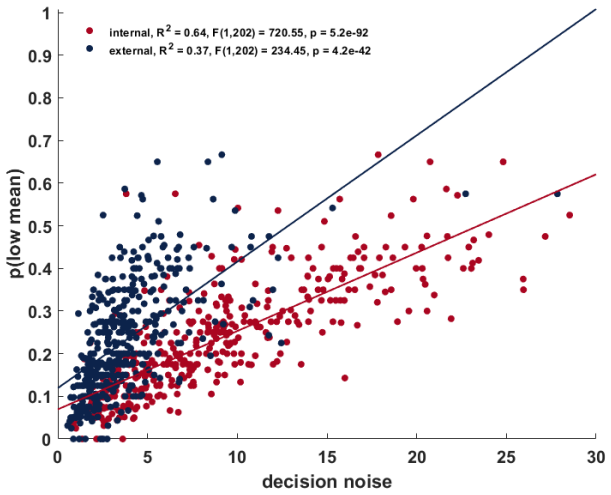


Figure 6: Both internal and external noise contribute to random exploration, internal noise is dominant.

Discussion

In this paper, we investigated whether random exploration is driven by internal noise, putatively arising in the brain, or external noise, arising from the environment. We find horizon dependent changes in both internal and external sources of noise, but that the effect of internal noise is much greater.

One limitation of this work is in the interpretation of the different noises as being internal and external. In particular, while we controlled many aspects of the stimulus across repeated games (e.g. the outcomes and the order of the forced trials), we could not perfectly control all stimuli the participant received which, for example, would vary based on where they were looking. Thus, our estimate of external noise is likely a lower bound. Likewise, our estimate of internal noise is likely an upper bound as these 'missing' sources of stimulus driven noise would be interpreted as internal noise in our model. Despite this, the sheer magnitude of the difference between internal and external noise (internal noise is 2-3 times the size of external noise, Figure XXX), suggests that our interpretation may be safe as an awful lot of external noise would have to be explained by variables not relevant to the task.

The horizon-dependent increase in internal noise is consistent with the idea that random exploration is driven by intrinsic variability in the brain. This is in line with work in the bird song literature in which song variability during song learning has been tied to neural variability arising from specific areas of the brain. In addition, this work is consistent with a recent report from Ebitz et al [7] in which the behavioral

variability of monkeys in an 'explore' state was also tied to internal rather than external sources of noise.

Whether such a noise-controlling area exists in the human brain is less well established, but one candidate theory [4] suggests that norepinephrine (NE) from the locus coeruleus may play a role in modulating internal levels of noise. While there is some evidence that NE plays a role in explore-exploit behavior [8], this link has been questioned [9].

MORE GENERALLY, OUT FINDING THAT INTERNAL DOMINATES BEHAVIORAL VARIABILITY OVER EXTERNAL NOISE, IS CONSISTENT WITH FINDINGS OF DRUGOWITSCH ET AL. IN PARTICULAR THESE AUTHORS SHOW THAT LINK IN THE Jan Drugowitsch stuff on variability. ALEX POUGET STUFF. INTERNAL NOISE COMES FROM COMPUTATIONAL ERRORS IN COMPUTING THE CORRECT STRATEGY IN LONG HORIZON CONDITIONS.

THREE IDEAS NOISE ADDED TO THE PROCESS ATTENTION = DECREASE SIGNAL VS INCREASE NOISE NOT NOISY BUT WRONG PLANNING WITH FEW SAMPLES DEEP EXPLORATION

Methods

Participants

A total of 204 participants from the UA subject pool did the task in 4 different experiments.

017

49 out of 148 participants are included for analysis. 59 were excluded because of the passive condition. 59 were excluded on the basis of performance (using the same exclusion criterion as in [2]) leaving 49 for the analysis.

075

58 out of 78 participants are included for analysis. 3 were excluded because of young age, 17 were excluded on the basis of performance (using the same exclusion criterion as in [2]) leaving 49 for the analysis.

personality

65 out of 91 participants are included for analysis. 26 were excluded on the basis of performance (using the same exclusion criterion as in [2]) leaving 65 for the analysis.

test-retest

33 out of 45 participants are included for analysis. 12 were excluded on the basis of performance (using the same exclusion criterion as in [2]) leaving 33 for the analysis.

Task

The task was a modified version of the Horizon Task (). As in the original paper, the distributions of payoffs tied to bandits were independent between games and drawn from a Gaussian distribution with variable means and fixed standard deviation of 8 points. Participants were informed that in every game one of the bandits always has a higher mean reward than the other. Differences between the mean payouts of the two slot machines were set to either 4, 8, 12 or 20. One of the means was always equal to either 40 or 60 and the second was set accordingly. The order of games was randomized. Mean sizes and order of presentation were counterbalanced. The number of games participants play depends on how well they perform, the better they perform, the sooner the task will end. On average, participants played 128 games and the whole task lasted between 39 and 50 minutes (mean 43.4 minutes).

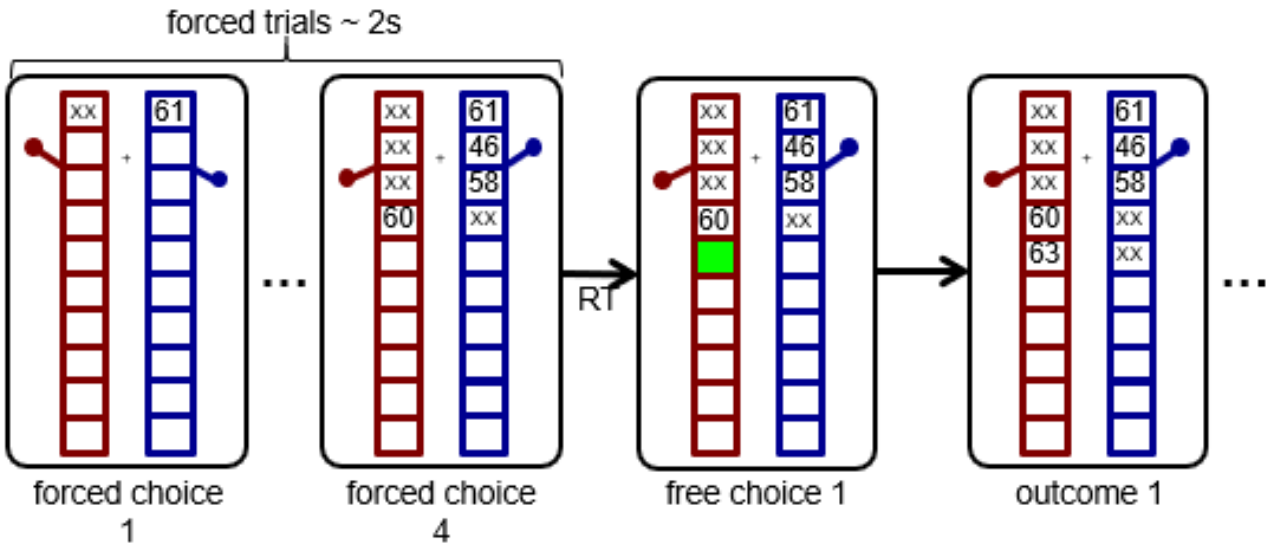


Figure 7: Timeline of a game

Each game consisted of 5 or 10 choices. Every game started with a fixation cross, then a bar of boxes will show up indicating the horizon for that game. For the first 4 games - the instructed games, we

highlight the box on one of the bandits to instruct the participant to choose that option, they have to press the corresponding key to reveal the outcome. From the 5^{th} trial, boxes on both bandits will be highlighted and they are free to make their own decision. There was no time limit for decisions. During free choices they could press either the left arrow key or right arrow key to indicate their choice of left or right bandit. The score feedback was presented for 500ms. The task was programmed using Psychtoolbox in MATLAB. ().

Data and code

Model-based analysis

We modeled behavior on the first free choice of the Horizon Task using a version of the logistic choice model in () that was modified to differentiate internal and external noises. In particular, we assume that in repeated games, external noise remains the same whereas internal noise can change.

Model Fitting

Hierarchical Bayesian Model

Each subject's behavior is described by 4 free parameters. These parameters are: the information bonus, A, in both horizon conditions, the spatial bias, B, in the four horizon x uncertainty conditions, and external decision noise, n_{int}, and internal decision noise, n_{ext} in the four horizon x uncertainty conditions (Table 2, Figure 7).

Each of the free parameters is fit to the behavior of each subject using a hierarchical Bayesian approach (). In this approach to model fitting, each parameter for each subject is assumed to be sampled from a group-level prior distribution whose parameters, the so-called 'hyperparameters', are estimated using a Markov Chain Monte Carlo (MCMC) sampling procedure. The hyper-parameters themselves are assumed to be sampled from 'hyperprior' distributions whose parameters are defined such that these hyperpriors are broad.

The particular priors and hyperpriors for each parameter are shown in Table 2. For example, we assume that the information bonus, A^{is}, for each information condition x horizon condition i, is sampled from a Gaussian prior with mean μ_i^A and standard deviation σ_i^A. These prior parameters are sampled in turn from their respective hyperpriors: μ_i^A, from a Gaussian distribution with mean 0 and standard deviation 10, σ_i^A

from an Exponential distribution with parameters 0.1.

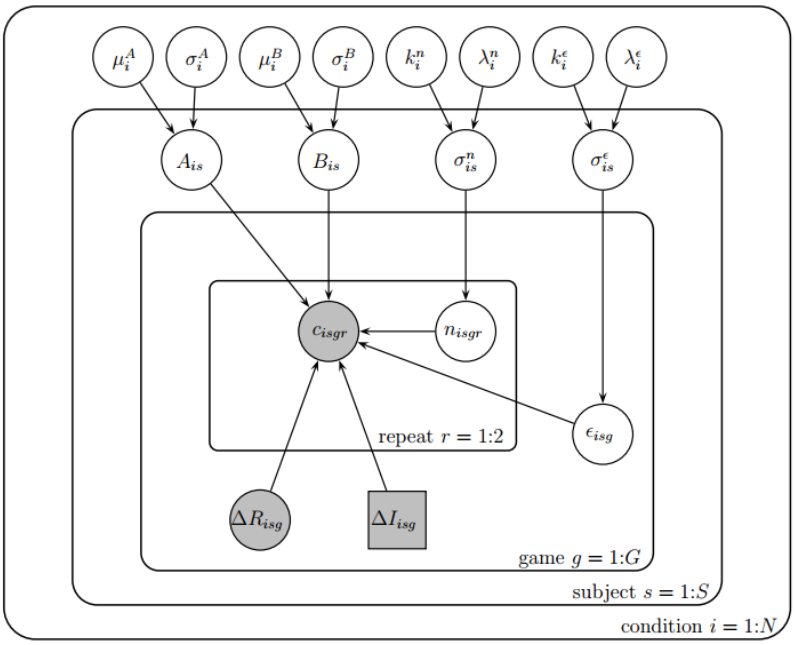
Parameter	Prior	Hyperparameters	Hyperpriors
information bonus, A_{is}	$A_{is} \sim$ Gaussian(μ_i^A, σ_i^A)	$\theta_i^A = (\mu_i^A, \sigma_i^A)$	$\mu_i^A \sim$ Gaussian(0, 100) $\sigma_i^A \sim$ Exponential(0.1)
spatial bias, B_{is}	$B_{is} \sim$ Gaussian(μ_i^B, σ_i^B)	$\theta_i^B = (\mu_i^B, \sigma_i^B)$	$\mu_i^B \sim$ Gaussian(0, 100) $\sigma_i^B \sim$ Exponential(0.1)
external decision noise, ϵ_{isg}	$\epsilon_{isg} \sim$ Gaussian(0, σ_{is}^{ext})	$\theta_i^{ext} = (k_i^{ext}, \lambda_i^{ext})$	$k_i^{ext} \sim$ Gaussian(0.1) $\lambda_i^{ext} \sim$ Exponential(10)
internal decision noise, σ_{isgr}	$\sigma_{isg} \sim$ Gaussian(0, σ_{is}^{int})	$\theta_i^{int} = (k_i^{int}, \lambda_i^{int})$	$k_i^{int} \sim$ Gaussian(0.1) $\lambda_i^{int} \sim$ Exponential(10)

Table 2: Model parameters, priors, hyperparameters and hyperpriors.

Model fitting using MCMC

The model was fit to the data using Markov Chain Monte Carlo approach implemented in the JAGS package () via the MATJAGS interface (psiexp.ss.uci.edu/research/programs_data/jags/). This package approximates the posterior distribution over model parameters by generating samples from this posterior distribution given the observed behavioral data.

In particular we used 4 independent Markov chains to generate 40000 samples from the posterior distribution over parameters (10000 samples per chain). Each chain had a burn in period of 1000 samples, which were discarded to reduce the effects of initial conditions, and posterior samples were acquired at a thin rate of 1. Convergence of the Markov chains was confirmed post hoc by eye. Code and data to replicate our analysis and reproduce our Figures is provided as part of the Supplementary Materials.



Priors

$\mu_i^A \sim \text{Gaussian}(0, 100)$, $\sigma_i^A \sim \text{Exponential}(0.1)$

$\mu_i^B \sim \text{Gaussian}(0, 100)$, $\sigma_i^B \sim \text{Exponential}(0.1)$

$k_i^n \sim \text{Exponential}(0.1)$, $\lambda_i^n \sim \text{Exponential}(10)$

$k_i^\epsilon \sim \text{Exponential}(0.1)$, $\lambda_i^\epsilon \sim \text{Exponential}(10)$

Subject specific parameters

$A_{is} \sim \text{Gaussian}(\mu_i^A, \sigma_i^A)$

$B_{is} \sim \text{Gaussian}(\mu_i^B, \sigma_i^B)$

$\sigma_{is}^n \sim \text{Gamma}(k_i^n, \lambda_i^n)$

$\sigma_{is}^\epsilon \sim \text{Gamma}(k_i^\epsilon, \lambda_i^\epsilon)$

Stimulus driven noise for repeated game

$\epsilon_{isg} \sim \text{Gaussian}(0, \sigma_{is}^\epsilon)$

Free noise for each game

$n_{isgr} \sim \text{Gaussian}(0, \sigma_{is}^n)$

Observed choices

$\Delta Q_{isgr} \leftarrow \Delta R_{isg} + A_{is}\Delta I_{isg} + B_{is} + n_{isgr} + \epsilon_{isg}$

$c_{isgr} \sim \text{Bernoulli}\,(Q_{isgr} > 0)$

Figure 8: Hierarchical Bayesian model

References