Separating random and deterministic sources of computational noise in explore-exploit decisions

[siyu 1] (was): The nature of decision noise in random exploration

Siyu Wang[1,¤] and Robert C. Wilson[1,2,3]

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

[2]Neuroscience and Physiological Sciences Graduate Interdisciplinary Program, University of Arizona, Tucson AZ, USA

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

[¤]Current Address: Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda MD, USA

December 18, 2023

Abstract

Human decision making is inherently variable. While this variability is often seen as a sign of suboptimal behavior, recent work suggests that variability can actually be adaptive. An example arises when we must choose between exploring unknown options or exploiting options we know well. A little randomness in these 'explore-exploit' decisions is remarkably effective as it can encourage us to explore options we might otherwise ignore. In line with this idea, several studies have found evidence that people increase their behavioral variability when it is valuable to explore. A key question, however, is whether this variability in so-called 'random exploration' is actually random.[bob] That is, is random exploration driven by stochastic processes in the brain or by some unobserved deterministic process that we have failed to account for when measuring behavioral variability? By designing an explore-exploit task in which, unbeknownst to them, participants are presented with the exact same choice twice, we provide a partial answer to this question. By modeling behavior in this task, we were able to estimate a lower bound on the amount of variability that is deterministically driven by the stimulus and an upper bound on the amount of variability that is random.[bob] Using this approach, we find evidence that at least 15%[siyu] of the variability in random exploration can be accounted for by deterministic processing of the stimulus. Conversely, this suggests that up to 85% of the variability is truly 'random,' although it is still possible that this variability is driven by deterministic factors not related to the stimulus.[bob] Finally, our results suggest that both deterministic and random sources of variability change proportionally to each other as the value of exploration increases, suggesting that a common noise gating mechanism may be at play in random exploration.[bob]

Introduction

Imagine trying to decide where to go to dinner on a date. You can go to your favorite restaurant, the one you both really enjoy and always go to, or you can try a new restaurant that you know nothing about. Such decisions, in which we must choose between a well-known 'exploit' option and a lesser known 'explore' option, are known as explore-exploit decisions. From a theoretical perspective, making optimal explore-exploit choices, i.e. choices that maximize long-term reward, is computationally intractable in most cases (**??**). In part because of this computational complexity, there is considerable interest in how humans and animals solve the explore-exploit dilemma in practice (**???**)[siyu].

One particularly effective strategy for solving the explore-exploit dilemma is choice randomization (**???**) , also known as random exploration[siyu]. In this strategy, high value 'exploit' options are not always chosen and exploratory choices are sometimes made by chance. In modeling terms, random exploration works by adding 'decision noise' to the value of the options such that sub-optimal exploratory options can sometimes have a higher total score (i.e., value + noise) than the exploit option and get chosen.[siyu] Such random exploration, is surprisingly effective and, if implemented correctly, can come close to optimal performance in some cases[siyu](**????**).

It has recently been shown that humans appear to use random exploration and can increase decision noise when it is more beneficial to explore (**??**) ALSO CITE FINDLING ET AL.[bob]. In one of these tasks, known as the Horizon Task (**?**), the key manipulation is the horizon condition, i.e. the number of decisions remaining for the participant to make. Increasing the horizon makes exploration more valuable as there is more time to use the information gained by exploration to maximize future rewards. For example, if you are leaving town tomorrow (short horizon), you will probably exploit the restaurant you know and love, but if you are in town for a while (long horizon), you will be more likely to explore the new restaurant. Using such a horizon manipulation it has been shown that people's behavior is more variable in long horizons than short horizons, suggesting that they use adaptive decision noise to solve the explore-exploit dilemma (**?**).

One limitation of this previous research, however, is that it is difficult to tell whether what we have called 'decision noise' comes from a truly random process. That is, whether behavioral variability is due to genuinely stochastic processes in the brain or whether it is due to deterministic processes that we failed to observe. Decision noise as defined in previous research essentially quantifies the extent to which behavior cannot be predicted by a computational model. A missing deterministic component from the

[siyu 2] (deleted): the decision process between exploration and exploitation is corrupted by [...]

[siyu 3] (was): In theory, such

[siyu 4] (was): One limitation of this previous work, is that it is [...]

[siyu 5] (was): simply

model could give rise to variability in behavior that might appear to be random noise.[bob] For example, in the restaurant example, my usual preference for one restaurant or another may be overruled if I see an ex romantic partner going into one of them. Avoiding an ex is a deterministic process, but if we fail to take the ex's presence into account as scientists modeling the decision, then over a series of such decisions where the ex is present or not, we would mistakenly attribute the ensuing 'variability' in choice to randomness.

In this paper, we investigate the extent to which the apparent randomness in random exploration can be explained by deterministic processing of the stimulus (which we refer to as 'deterministic noise') vs other processes, including deterministic processing of unobserved stimuli as well as truly stochastic processes (which we refer to as 'random noise')[bob]. To distinguish between these two types of noise[siyu] we modify the Horizon Task (**?**) to have people face the exact same explore-exploit choice twice. If the decision is a purely deterministic function of the stimulus (i.e., decision noise is purely deterministic noise), then people's choices should be identical for both decisions, since the stimulus is the same both times. Conversely, if the decision is a purely random function of the stimulus (i.e., decision noise is purely random noise), then people's choices will be different 50% of the time, since the random noise is different each time. In between these two extremes of purely deterministic and purely random drivers of behavioral variability, the extent to which people's decisions are consistent between the two decisions can be used to estimate the amount of deterministic and random noise.[bob]

In the following, we analyze behavior on the repeated decisions version of the Horizon Task in both a model-free and model-based manner. Our model-free analysis estimates the extent to which people's behavior is consistent across repeated versions of the same decision. By measuring how this choice consistency changes as a function of horizon, this model-free analysis offers qualitative insight into the extent to which behavioral variability is driven by deterministic vs random noise. Our model-based analysis uses a computational model of the explore-exploit decision in the Horizon Task that incorporates both noise processes. By fitting this model to the behavioral data, this model-based analysis allows us to quantify the relative size of the two sources of noise and how they change in the service of exploration.[bob]

Results

The Repeated-Games Horizon Task

We used a modified version of the 'Horizon Task' (**?**) to show the influence of stimulus-driven 'deterministic noise' vs non-stimulus-driven 'random noise' in explore-exploit decisions (Figure 1). In this task, participants make a series of BE CAREFUL USING THE WORD REPEATED. ONLY USE IT TO REFER TO THE REPEATED GAMES - HERE I REPLACED REPEATED WITH 'SERIES OF'[bob] choices between two slot machines, or 'one-armed bandits,' that pay out probabilistic rewards. Because they are initially unsure as to the mean payoff of each bandit, this task requires that participants carefully balance exploration of the lesser known bandit with exploitation of the better known bandit to maximize their overall rewards.

[siyu 6] (was): on people's

THERE NEEDED TO BE MORE INFORMATION ABOUT THE HORIZON TASK HERE - REMEMBER THIS COUDL BE THE FIRST TIME PEOPLE HAVE SEEN THE TASK. The Horizon Task has two key features that together allow it to quantify explore-exploit behavior. The first of these features is the time horizon — the number of decisions participants will make in the future. By changing this horizon from short (1 trial) to long (6 trials), the Horizon Task allows us to control the relative value of exploration and exploitation. Just like the restaurant example in the introduction, when the horizon is short, participants should be more likely to exploit the option they believe to be best, because this leads to the highest payoff in the short term. Conversely, when the horizon is long, participants should be more likely to explore at first, because this allows them to gather information to make better choices later on. By contrasting behavior between short and long horizon conditions *on the very first trial*, when all else is equal, the Horizon Task allows us to quantify how behavior changes when it is valuable to explore.

The second key feature of the Horizon Task are the four 'example plays' at the start of each game that allow us to control exactly what participants know about the two bandits before they make their choice. In these example plays, participants are instructed which of the bandits to play allowing us to control how much information they have about each of the options. The instructed trials are used to set up one of two information conditions and 'unequal information' or [1 3] condition, in which participants play one option once and the other three times, and an 'equal information' or [2 2] condition, in which participants play both options twice.[bob]

The second key feature of the Horizon Task are the four 'example plays' at the start of each game that allow us to control exactly what participants know about the two bandits before they make their choice.

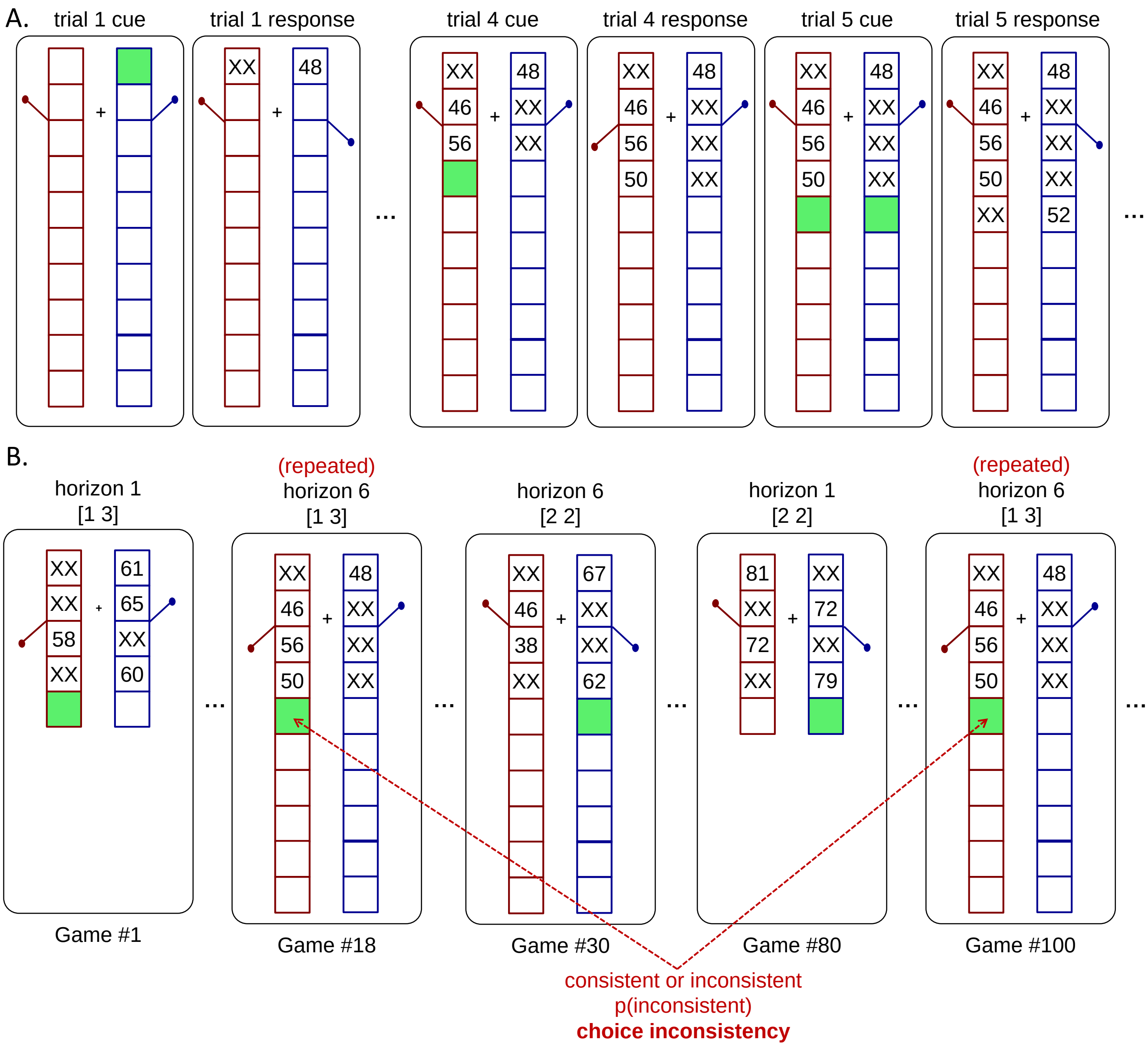


Figure 1: Schematic of the experiment. (A) Dynamics of an example horizon 6 game. Here the first four trials are forced trials in which participants are instructed which option to play. After the forced trials, participants are free to choose between the two options for the remainder of the game. (B) Example repeated games over the course of the experiment. On average, participants play more than 150 such games, with varying horizon (1 vs 6), uncertainty condition ([1 3] vs [2 2]) and observed rewards. In addition, all games are repeated (as Game 18 and 100 are here) such that participants will be faced with the exact same pattern of forced trials and exact same outcomes from those forced trials twice within each experiment. These repeated games allow us to compute the relative contribution of deterministic and random noise by analyzing the extent to which choices are *consistent* across the repeated games.

In these example plays, participants are instructed which of the bandits to play allowing us to control how much information they have about each of the options. The instructed trials are used to set up one of two information conditions and 'unequal information' or [1 3] condition, in which participants play one option once and the other three times, and an 'equal information' or [2 2] condition, in which participants play both options twice.

Relative to the original Horizon Task, the key modification in this paper is to give people 'repeated games,' in which they see exact same set of example plays twice in two separate games separated by several minutes in time so as to avoid detection. By repeating the instructed plays for each game twice, we can set up a situation where (unbeknownst to the participants) they are faced with the exact same explore-exploit choice, with the exact same stimuli twice. Thus, if their behavior is a deterministic function of the stimuli, then they will make the same decision in both games and their choices will be consistent. Conversely, if their behavior is not driven by a deterministic function of the stimulus, then their choices on the repeated games will be inconsistent some fraction of the time.[bob] The extent to which participants' choices are consistent on the repeated versions of the games allow us to quantify the extent to which the variability in[siyu] their behavior was driven by a deterministic process vs a random noise process.

[siyu 7] (was): On average participants played 67.31 repeated games each allowing

[siyu 8] (was): a deterministic function of the stimulus or not

Both behavioral variability and information seeking increase with horizon

GENERAL COMMENT - WE NEED TO WALK PEOPLE THROUGH THE BASIC RESULTS MUCH MORE SLOWLY, OTHERWISE WILL BE CONFUSING FOR PEOPLE NEW TO THE HORIZON TASK.[bob]

Before discussing the results for repeated games, we first confirm that the basic behavior in this task is consistent with our previously reported results using both a model-free and model-based approach (**?**). In both analyses, we focus on just the first free-choice trial in each game, where the only thing that differs between the horizon conditions is the number of choices that participants will make in the future.[bob]

Model-free analysis

In the model-free analysis, we quantify random and directed exploration using simple choice probabilities. Random exploration is quantified as the probability of choosing the option that has the lower average payout in the example plays in the equal, or [2 2], condition, $p(\text{low mean})$. The idea here is that, in the equal condition, the optimal strategy is to compute the mean payout for each bandit from the example plays and then always choose the option with the highest mean. When participants do not choose the option with the higher mean, the assumption is that this is due to some kind of 'decision noise,' making the probability of choosing the low mean option a measure of behavioral variability. In this view, random exploration corresponds to an increase in $p(\text{low mean})$ with horizon, which is exactly what we see in the data (Figure 2A; t(64) = 7.99, p $<$ 0.001 for [2 2]).[bob]

Directed exploration is quantified as the probability of choosing the more informative option $p(\text{high info})$ in the unequal, or [1 3], condition. The more informative option is the option played once during the example plays as choosing this option gives relatively more information (doubling the number of samples from 1 to 2) than choosing the option played three times (only increasing the number of sample by a third, from 3 to 4). In this view, directed exporation corresponds to an increase in $p(\text{high info})$ with horizon, which is exactly what we see in the data (Figure 2B; t(64) = 6.92, p $<$ 0.001).[bob]

Model-based analysis

I'M ADDING THIS SECTION, BECAUSE IT ALLOWS US TO INTRODUCE THE MODEL EARLY. THAT WAY WHEN THE MODEL COMES IN LATER WE CAN SEE HOW IT'S CHANGED[bob]

Another approach to understanding behavior in the Horizon Task is to use a computational model. In this case, we model participants' choices on the first free-choice trial by assuming they make decisions by computing the difference in value (or utility) ΔQ between the right and left options, choosing right when $\Delta Q > 0$ and left otherwise. Specifically, we write

$$\Delta Q = \Delta R + A\Delta I + b + n \tag{1}$$

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 in 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 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_{det} denotes the deterministic noise, which is identical on the repeat versions of each game; and n denotes decision noise, which, in this version of the model is a combination of deterministic and random noise. n is assumed to come from a logistic distributions with mean 0 and standard deviations σ. [bob]

The free parameters of this model are: the information bonus A, which controls the level of directed exploration; the noise standard deviation, σ, which controls the level of random exploration, and the spatial bias, b, which determines the extent to which participants prefer the option on the right. These free parameters are fit separately for each participant in each horizon and information condition, allowing us to test whether directed and random exploration increase with horizon. Consistent with previous research, we find that this is indeed the case (Figure 2C, D; STATS FOR CHANGE IN INFO BONUS AND DECISION NOISE). [bob]

Taken together our model-free and model-based analyses agree with previous findings showing increased behavioral variability and increased information seeking in the long horizon condition consistent with humans using random and directed exploration. However, for random exploration, this previous analysis cannot distinguish between deterministic and random sources of noise. For this we analyze the extent to which people's choices are consistent on the repeated games.[bob]

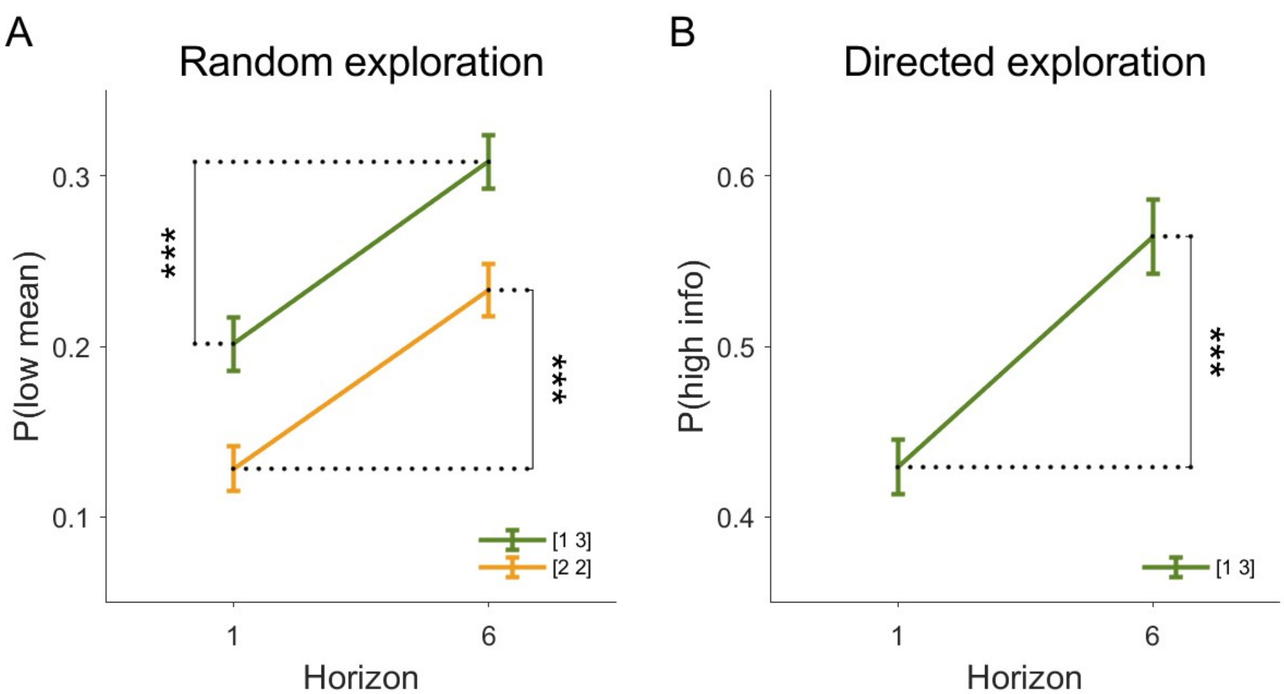


Figure 2: ADD C AND D PANELS FOR MODEL-BASED RESULTS HERE[bob] Replication of previous findings. Both p(low mean) (A) and p(high info) (B) increase with horizon suggesting that people use both random and directed exploration in this task.

Model-free analysis of repeated games suggests that random exploration involves both random and deterministic noise

Next we asked whether participants' choices were consistent or inconsistent in the two repetitions of each game. The idea behind this measure is that purely deterministic noise should lead to consistent choices as the deterministic stimulus is identical both times. Conversely, if choice is not entirely driven by a deterministic process and is also driven by random noise, participants' choices should be more inconsistent across the repetitions of the game. Moreover, if decision noise is purely random noise, meaning there is no unobserved deterministic process, we will show that we can actually predict the expected level of choice inconsistencies across repetitions of games by accounting for the known deterministic processes and assuming that the random noise process is independent in repetitions of the game.[siyu]

[siyu 9] (deleted): Conversely, if choice is not a deterministic function of the [...]

To quantify choice inconsistency we computed the frequency with which participants made different responses for pairs of repeated games (Figure 3, Supplementary Figure S2). Using this measure we found that participants made inconsistent choices in both the unequal ([1 3]) and equal ([2 2]) information conditions, suggesting that not all of the noise was stimulus driven (t-test vs zero revealed that inconsistency was greater than zero for all horizon and uncertainty conditions). In addition, we found that choice inconsistency was higher in horizon 6 than in horizon 1 for both [1 3] and [2 2] condition (For [1 3] condition, $t(64) = 5.41$, $p < 0.001$; for [2 2] condition, $t(64) = 6.26$, $p < 0.001$), suggesting that at least some of the horizon dependent noise is not a deterministic function of the stimulus, but rather random noise[siyu].

[siyu 10] (deleted): For [1 3] condition, $t(64) = 13.72$, $p < 0.001$ [...]

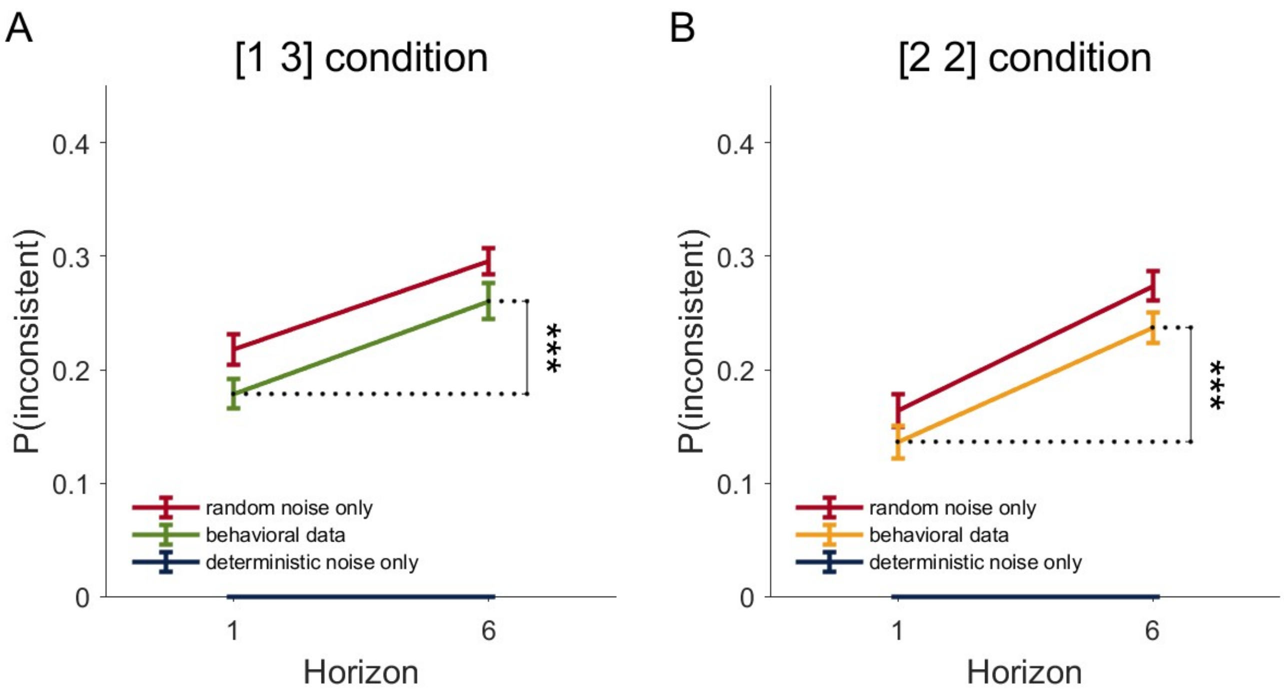


Figure 3: Model-free analysis suggests that both deterministic and random noise contribute to the choice variability in random exploration. For both the [1 3] (A) and [2 2] (B) condition, people show greater choice inconsistency in horizon 6 than horizon 1. However, the extent to which their choices are inconsistent lies between what is predicted by purely deterministic and random noise, suggesting that both noise sources influence the decision.

To gain more quantitative insight into these results, we computed theoretical values for the choice inconsistency for the purely deterministic and purely random noise cases. For purely deterministic noise this computation is simple because people should make the exact same decisions each time in repeated games, meaning that $p(\text{inconsistent}) = 0$ in this case. For purely random noise, the two games should be treated independently, allowing us to compute the choice inconsistency in terms of the probability of choosing the low mean option, $p(\text{low mean})$, as

$$\begin{aligned} p(\text{consistent}) &= p(\text{low mean})^2 + p(\text{high mean})^2 \\ &= p(\text{low mean})^2 + (1 - p(\text{low mean}))^2 \\ \text{hence,} \quad p(\text{inconsistent}) &= 1 - p(\text{consistent}) = 2p(\text{low mean})(1 - p(\text{low mean})) \end{aligned}$$

Furthermore, to account for the fact that $p(\text{low mean})$ is a function of reward difference ΔR between the two bandits and the information condition I, we estimated the conditional probability:

$$p(\text{inconsistent}|\Delta R, I) = 2p(\text{low mean}|\Delta R, I)(1 - p(\text{low mean}|\Delta R, I))$$

Then based on the likelihood that each condition (ΔR vs I) occurs in the task $\rho(\Delta R, I)$, we have

$$p(\text{inconsistent}) = \sum_{\Delta R, I} \rho(\Delta R, I) p(\text{inconsistent}|\Delta R, I)$$

siyu

As shown in Figure 3, people's behavior falls in between the pure deterministic noise prediction and the pure random noise prediction. Specifically, behavior is different from the pure random noise prediction in the both the [1 3] condition (t(64) = 4.83, $p < 0.001$ for horizon 1, t(64) = 3.12 $p = 0.003$ for horizon 6) and the [2 2] condition (t(64) = 3.92, $p < 0.001$ for horizon 1, t(64) = 3.71, $p < 0.001$ for horizon 6). Likewise, behavior is different from pure deterministic noise prediction in both the [1 3] condition (t(64) = 13.72, $p < 0.001$ for horizon 1, t(64) = 16.71, $p < 0.001$ for horizon 6) and the [2 2] condition (t(64) = 9.55, $p < 0.001$ for horizon 1, t(64) = 17.93, $p < 0.001$ for horizon 6). As a negative control of our method for estimating $p(\text{inconsistent})$ for purely random noise, we simulated choices using a decision model that only includes random noise (Equation 1), and found that $p(\text{inconsistent})$ in this simulated data is not different from our pure random noise prediction in all horizon and uncertainty conditions ($p > 0.05$, Supplementary Figure S3).[bob] Together, our results suggest that both random noise and deterministic noise contribute to the choice variability in random exploration.. However, how each of these types of noise change with horizon is difficult to discern.

[siyu 11] (was): This suggests that at least some of the 'noise' is [...]

Model-based analysis provides a lower-bound estimate of deterministic noise and an upper-bound estimate of random noise

[siyu 12] (was): shows deterministic noise changes with horizon

To more precisely quantify the contribution of deterministic noise and random noise[siyu], we turned to model fitting. We modeled behavior on the first free choice of the Horizon Task using a version of the logistic choice model in (**?**) (Equation 1) that was modified to differentiate between components of the noise that are deterministically driven by the stimulus ('deterministic noise') and components of the noise that are not deterministically driven by the stimulus ('random noise'). In particular, we assume that in repeated games, the value of stimulus-driven deterministic noise is frozen whereas random noise is drawn independently both times.

[siyu 13] (was): deterministic noise from random noise

Overview of model

To model participants' choices on the first free-choice trial, we use a modified version of Equation 1. [bob]

$$\Delta Q = \Delta R + A\Delta I + b + n_{det} + n_{ran} \quad (2)$$

where, as before ΔR, is the the difference in mean rewards shown on the forced trials, ΔI, is the difference in information, A is the information bonus, and b is the spatial bias. New in Equation 2 are the terms n_{det} and n_{ran}. n_{det} denotes the deterministic noise, which is identical on the repeat versions of each game; and n_{ran} denotes random noise, which is uncorrelated between repeated plays and changes every game. n_{det} and n_{ran} are assumed to come from logistic distributions with mean 0, and standard deviations σ_{det} and σ_{ran}.

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

We used hierarchical Bayesian analysis to fit the parameters of the model (see Figure 11 for a graphical representation of the model in the style of **?**). In particular, we fit values of the information bonus A, spatial bias b, variance of random noise σ^2_{ran}, and variance of deterministic noise, σ^2_{det} for each participant in each horizon. Model fitting was performed using the MATJAGS and JAGS software (**??**) with full details given in the Methods.

Model validation

[siyu 14] (was): Parameter recovery

To be sure that our fit parameter values were meaningful and to understand the limits of our model, we evaluated our model extensively using simulated data. This allowed us to quantify whether deterministic and random noise can be identified under ideal conditions where the behavior is generated by the model with known parameters. Full details of this analysis are presented in the Supplementary Materials.

In this section we focus on our results for parameter recovery. In a parameter recover analysis, behavioral data is simulated by the model with known parameters and then this simulated behavioral data fit with the model to quantify the extent to which fit parameters match the input simulated parameters — that is, whether the simulated parameters can be recovered.

Parameter recover in this task was good for this model, with fit values for σ_{ran} and σ_{det} showing strong correlations with their simulated values (INSERT STATS HERE). However, while the relationship was near perfect for random noise, there was a systematic bias to underestimate the level of deterministic noise by about XXX (INSERT VALUE HERE). Despite this underestimation of deterministic noise in both horizon conditions, the difference in deterministic noise between horizons is much better captured (STATS). This is because the underestimation of deterministic noise is partially canceled out when the difference is taken between horizon conditions.

INSERT A SIMPLE PARAMETER RECOVERY PLOT FOR SIGMARAN AND SIGMADET.

Thus, overall, we were able to detect both deterministic and random noises using our model. Our model provides a lower bound for deterministic noise and an upper bound for random noise. In addition, we see better parameter recovery for random noise than deterministic noise. This is likely because we effectively have half as many trials for deterministic noise. In particular, while we generate two samples of random noise for each repeated game pair, we only generate one sample of deterministic noise, which by definition is the same in both of the repeated games. [bob]

In this section we focus on our results for parameter recovery. In a parameter recover analysis, behavioral data is simulated by the model with known parameters and then this simulated behavioral data fit with the model to quantify the extent to which fit parameters match the input simulated parameters — that is, whether the simulated parameters can be recovered.

Parameter recover in this task was good for this model, with fit values for σ_{ran} and σ_{det} showing strong correlations with their simulated values (INSERT STATS HERE). However, while the relationship was near perfect for random noise, there was a systematic bias to underestimate the level of deterministic noise by about XXX (INSERT VALUE HERE). Despite this underestimation of deterministic noise in both horizon conditions, the difference in deterministic noise between horizons is much better captured (STATS). This is because the underestimation of deterministic noise is partially canceled out when the difference is taken between horizon conditions.

INSERT A SIMPLE PARAMETER RECOVERY PLOT FOR SIGMARAN AND SIGMADET.

Thus, overall, we were able to detect both deterministic and random noises using our model. Our model provides a lower bound for deterministic noise and an upper bound for random noise. In addition, we see better parameter recovery for random noise than deterministic noise. This is likely because we effectively have half as many trials for deterministic noise. In particular, while we generate two samples of random noise for each repeated game pair, we only generate one sample of deterministic noise, which by definition is the same in both of the repeated games.

MOVE THE REST OF THIS SECTION TO THE SUPPLEMENTARY MATERIAL Firstly, we checked if our fitted deterministic noise could indeed capture unobserved deterministic process that was not accounted for by the decision model. We test this by leaving out one known deterministic process from the decision model, and ask if our method could recover that known deterministic process as deterministic noise. In particular, we fit a reduced version of our model that only considers reward and ignores the influence of uncertainty condition on explore-exploit decisions.

$$\Delta Q = \Delta R + n_{det} + n_{ran}$$

Here, $\Delta Q, \Delta R, n_{det}, n_{ran}$ represent the same variables as in the full model. If deterministic noise in our model can indeed capture unobserved deterministic processes that's missed by the model, then we would expect to see a higher level of fitted deterministic noise in the reduced model compared to in the full model, whereas the level of random noise should remain unchanged. By comparing the fitted posterior distributions over the group-level means of the deterministic and random noise parameters σ_{det} and σ_{ran}, as expected, we observed an increase in deterministic noise and no change in random noise between the reduced and the full model (Figure 4). This suggests that our model is capable of detecting missing deterministic processes. siyu

In this section we focus on our results for parameter recovery. In a parameter recover analysis, behavioral data is simulated by the model with known parameters and then this simulated behavioral data fit with the model to quantify the extent to which fit parameters match the input simulated parameters — that is, whether the simulated parameters can be recovered.

Parameter recover in this task was good for this model, with fit values for σ_{ran} and σ_{det} showing strong correlations with their simulated values (INSERT STATS HERE). However, while the relationship was near perfect for random noise, there was a systematic bias to underestimate the level of deterministic noise by about XXX (INSERT VALUE HERE). Despite this underestimation of deterministic noise in both horizon conditions, the difference in deterministic noise between horizons is much better captured (STATS). This is because the underestimation of deterministic noise is partially canceled out when the difference is taken between horizon conditions.

INSERT A SIMPLE PARAMETER RECOVERY PLOT FOR SIGMARAN AND SIGMADET.

Thus, overall, we were able to detect both deterministic and random noises using our model. Our model provides a lower bound for deterministic noise and an upper bound for random noise. In addition, we see better parameter recovery for random noise than deterministic noise. This is likely because we effectively have half as many trials for deterministic noise. In particular, while we generate two samples of random noise for each repeated game pair, we only generate one sample of deterministic noise, which by definition is the same in both of the repeated games.

MOVE THE REST OF THIS SECTION TO THE SUPPLEMENTARY MATERIAL Firstly, we checked if our fitted deterministic noise could indeed capture unobserved deterministic process that was not accounted for by the decision model. We test this by leaving out one known deterministic process from the decision model, and ask if our method could recover that known deterministic process as deterministic noise. In particular, we fit a reduced version of our model that only considers reward and ignores the influence of uncertainty condition on explore-exploit decisions.

$$\Delta Q = \Delta R + n_{det} + n_{ran}$$

Here, $\Delta Q, \Delta R, n_{det}, n_{ran}$ represent the same variables as in the full model. If deterministic noise in our model can indeed capture unobserved deterministic processes that's missed by the model, then we would expect to see a higher level of fitted deterministic noise in the reduced model compared to in the full model, whereas the level of random noise should remain unchanged. By comparing the fitted posterior distributions over the group-level means of the deterministic and random noise parameters σ_{det} and σ_{ran}, as expected, we observed an increase in deterministic noise and no change in random noise between the reduced and the full model (Figure 4). This suggests that our model is capable of detecting missing deterministic processes.

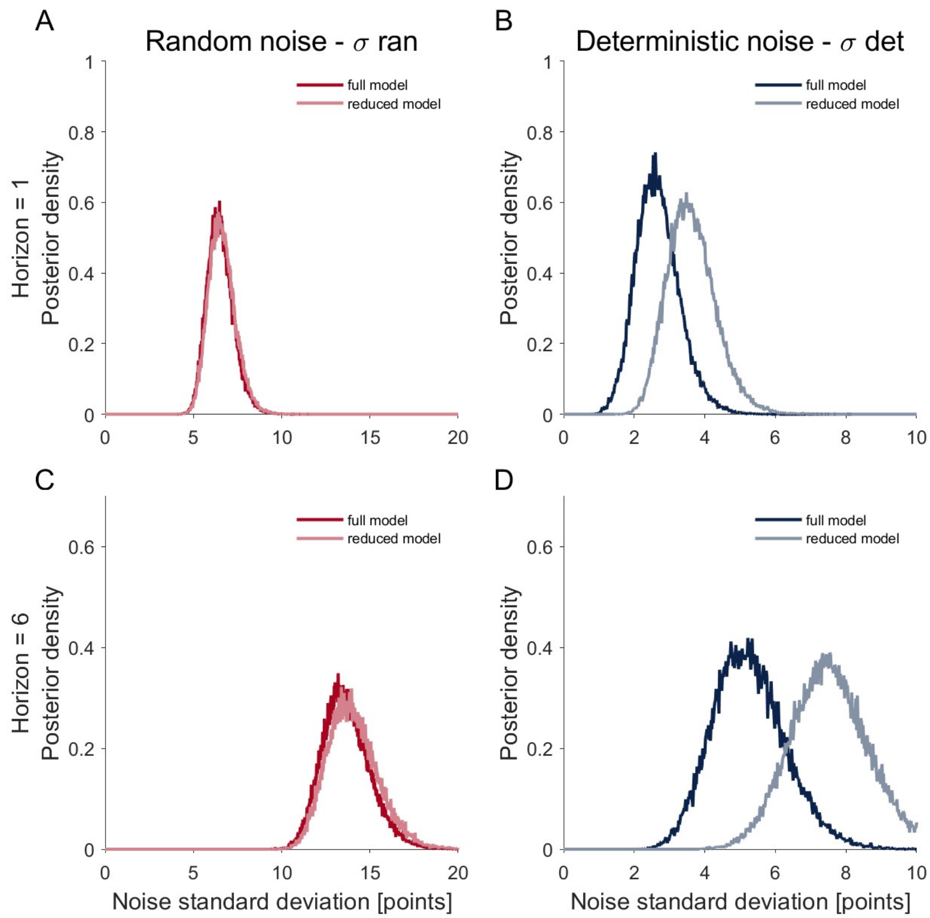


Figure 4: FONT SIZE ON LEGEND IS VERY SMALL![bob] Deterministic noise can recover known deterministic processes that's intentionally omitted by the model. In the reduced model where the deterministic effect of uncertainty condition is omitted from the model, deterministic noise is higher compared to the full model that accounts for the effect of uncertainty. Random noise remains unchanged between the two models.

Secondly, we evaluated our hierarchical Bayesian analysis procedure using the 'frequentist coverage analysis'. In the coverage analysis, we simulated choices with the fitted parameters from the Hierarchical Bayesian analysis, and then re-fit the simulated choices to see whether we can recover the parameters (Figure 5, Supplementary Figure S4). The simulation and re-fitting was repeated for 200 times. Then we counted out of the 200 repetitions how many times the true parameter that we simulate the choices from lies in the fitted 95% confidence interval. If our model fitting is reliable, then the fitted confidence interval should cover the true parameter for more than 95% of the simulations (this ratio will be referred to as the coverage rate). For random noise, the coverage rate is 100% for both horizon 1, horizon 6, and the horizon difference. For deterministic noise, the coverage rate is 66% for horizon 1 and 69% for horizon 6. By comparing the posterior distributions of parameters that were used to generate simulations and the posterior distribution of recovered parameters, it is clear that our model systematically underestimates deterministic noise (Figure 5). Despite the underestimation of deterministic noise in both horizons, we could still reliably detect the horizon changes of deterministic noise (coverage rate is 97%). This is because the underestimation of deterministic noise is partially canceled out when the difference is taken between horizons. For random noise, our model fitting procedure yields a faithful recovery. However, there is a conceptual limitation. Because random noise is modeled as non-stimulus-driven noise, it can include both true stochastic neural noise and possible deterministic noises which do not depend on the stimuli. Because of this, our random noise estimate provides an upper bound of true 'random noise' induced by intrinsic stochastic processes in the brain.[siyu]

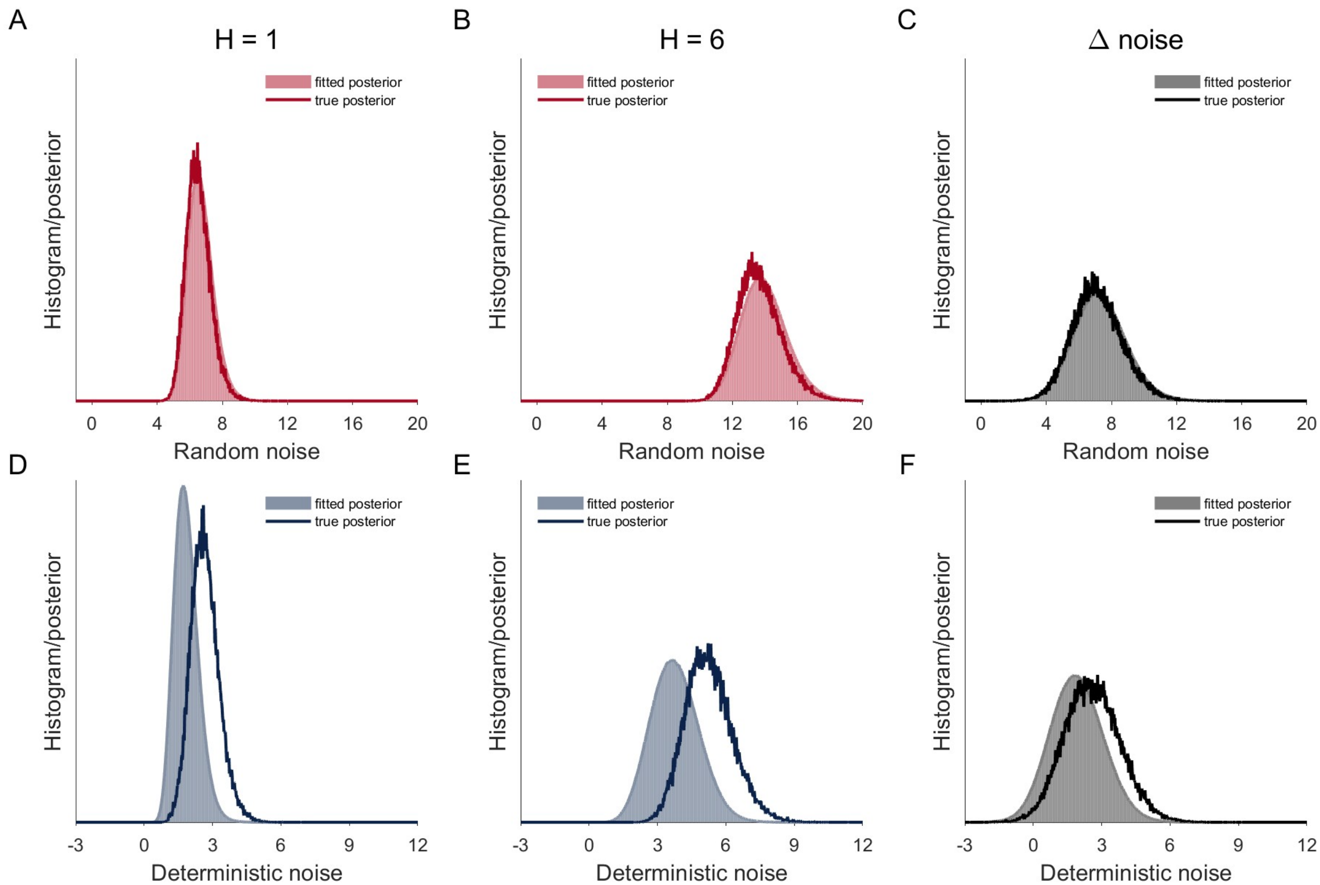


Figure 5: Parameter recovery over the posterior distribution of random and deterministic noise standard deviations σ_{det} and σ_{ran}. Solid lines are true posterior used to simulate choices. Lighter color shades represent the re-fitted posterior to the simulated choices. Our model fitting procedure faithfully recovers the non-stimulus-driven random noise (A, B), but systematically underestimates deterministic noise in both horizons (D, E). The horizon differences in random noise is also faithfully recovered (C). The horizon differences in deterministic noise is also underestimated but not significant (F).

Next, we tested the ability of our model fitting procedure to recovery parameters from simulated data at the subject level (Supplementary Figure S5 and S6). The correlations between the true vs fitted parameters are significant across participants for all parameters ($p < 0.001$). The strength of correlation between simulated and fit values are strong for both deterministic noise ($R > 0.8$) and random noise ($R > 0.9$). Despite the strong inter-subject correlations, we again observed a systematic underestimation of σ_{det} (Supplementary Figure S5 and S6).[siyu]

Lastly, in addition to testing how our model performs in parameter ranges around the actual fitted parameters, we tested the limitations of our models in arbitrary combinations of random vs deterministic noises. All combinations of random and deterministic noises with $0 \leq \sigma_{det} \leq 10$ and $0 \leq \sigma_{ran} \leq 10$ were tested. In a special case, we evaluated how our model performs when there is only random noise or only deterministic noise (Figure 6). In the simulation with fully deterministic noise and 0 random noise, our model successfully recovered both random and deterministic noise (Figure 6 C, D), however in the simulation with fully random noise and 0 deterministic noise, although our model successfully recovered random noise, some small proportion of deterministic noise was falsely detected when they should instead be 0 (Figure 6 A, B). However, this phenomenon only exists when the true deterministic noise is 0, once the true deterministic noise is greater than 1, we don't observe this inflation of deterministic noise anymore (Supplementary Figures S7). Apart from this, our model did a fairly good job in recovering all combinations of random and deterministic noises (Supplementary Figures S7).[siyu]

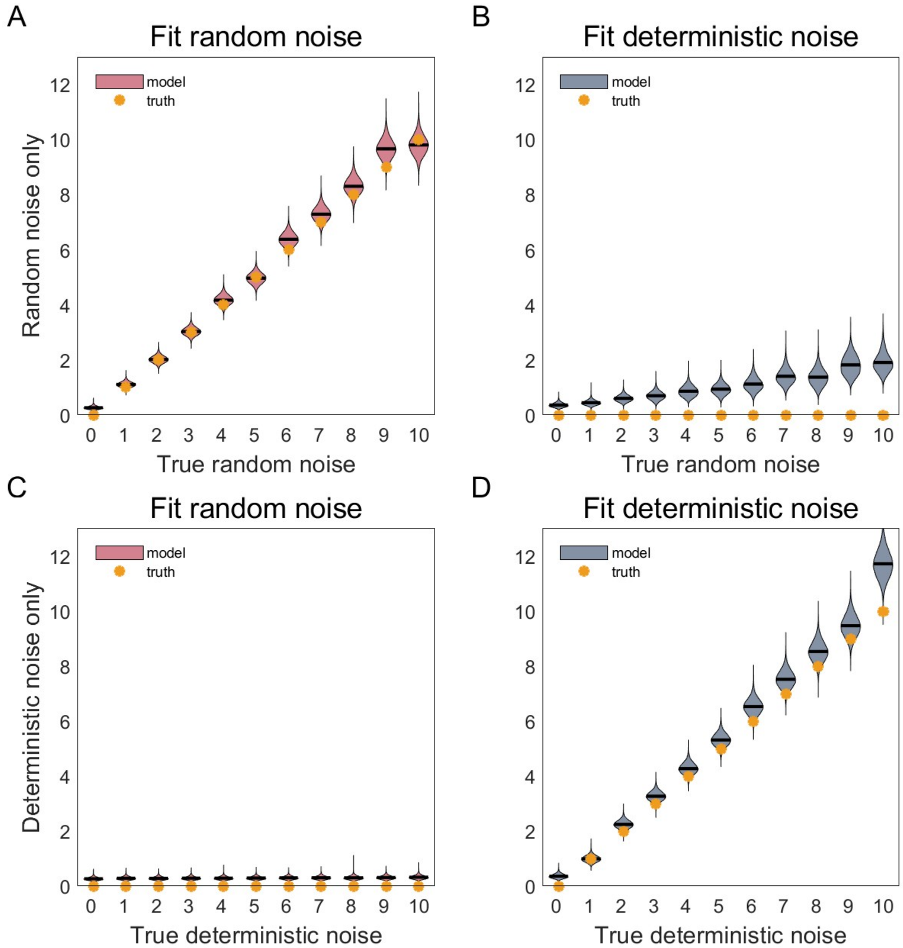


Figure 6: Parameter recovery over the posterior of random noise standard deviation, σ_{ran}, and deterministic noise standard deviation, σ_{det}, for purely random noise (top row) and purely deterministic noise (bottom row) games.

Overall, we are able to detect both deterministic and random noises using our model to a satisfactory extent. Our model provides a lower bound for deterministic noise and an upper bound for random noise. In addition, We see better parameter recovery for random noise than deterministic noise. This is likely because we effectively have half as many trials for deterministic noise. In particular, while we generate two samples of random noise for each repeated game pair, we only generate one sample of deterministic noise, which by definition is the same in both of the repeated games. [siyu]

Model-based results

Posterior distributions over the group-level means of the deterministic and random noise standard deviation σ_{det} and σ_{ran} are shown in Figure 7 and Supplementary Figure S8. Consistent with our model-free results, we see that both random and deterministic noise are non-zero. Numerically, random noise is about 2-3 times larger than the deterministic noise. By computing the posterior distribution of $\sigma^2_{det}/(\sigma^2_{det}+\sigma^2_{ran})$, our data suggests that 14.25% of the variability in random exploration is accounted for by deterministic noise ([4.90%, 28.81%], 95% CI).[siyu] In addition, we find that both random and deterministic noise increase with horizon. This increase was larger for random noise (mean = 7.13, 100% of samples showed an increase in random noise with horizon) than deterministic noise (mean = 2.59, 98.64% of samples showed an increase in deterministic noise with horizon). But intriguingly, the relative increase in both types of noise was similar (Figure 8). That is, when we compute the relative increase in deterministic noise with horizon, $\sigma^{det}_{horizon6}/\sigma^{det}_{horizon1}$, it is very similar to the relative increase in random noise with horizon $\sigma^{ran}_{horizon6}/\sigma^{ran}_{horizon1}$.

[siyu 15] (was): and that

[siyu 16] (was): the same

[siyu 17] (was): almost identical

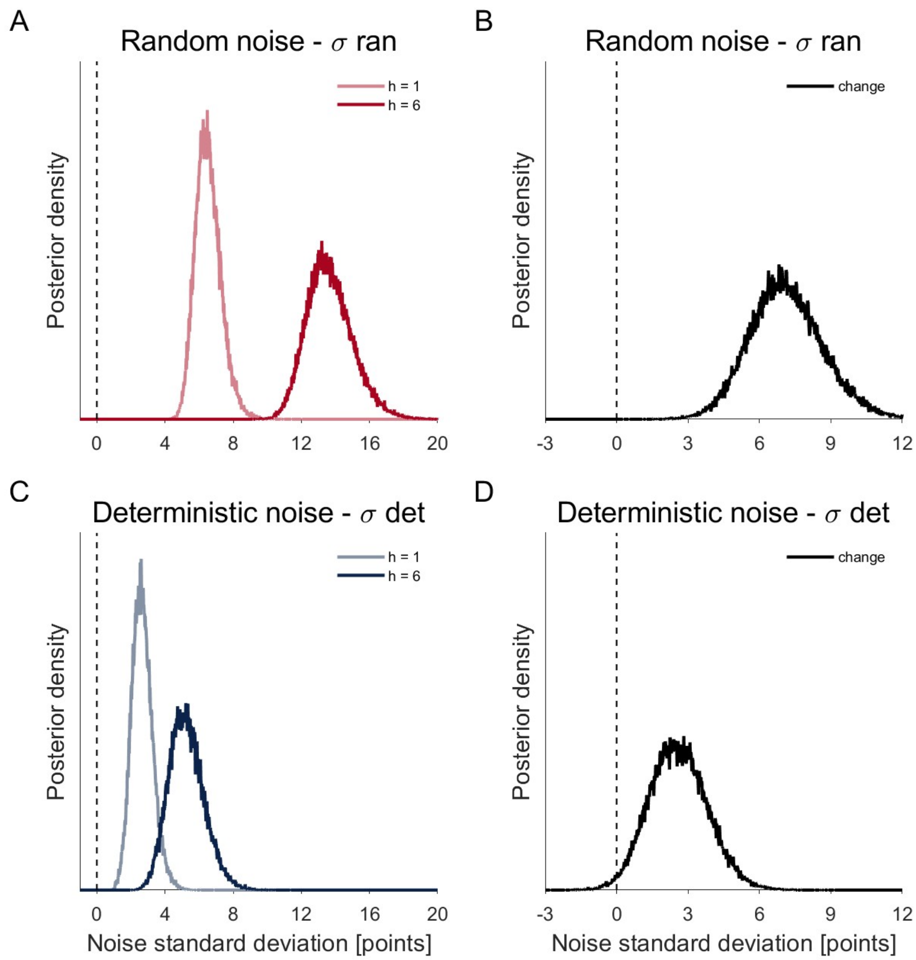


Figure 7: TITLE ON B AND D SHOULD SAY 'CHANGE IN RANDOM NOISE'[bob]Model based analysis showing the posterior distributions over the group-level mean of the standard deviations of random and deterministic noise. Both random (A, B) and deterministic (C,D) noises are nonzero (A, C) and change with horizon (B, D). However, random noise has both a greater magnitude overall (A, C) and a greater change with horizon (B, D) than deterministic noise.

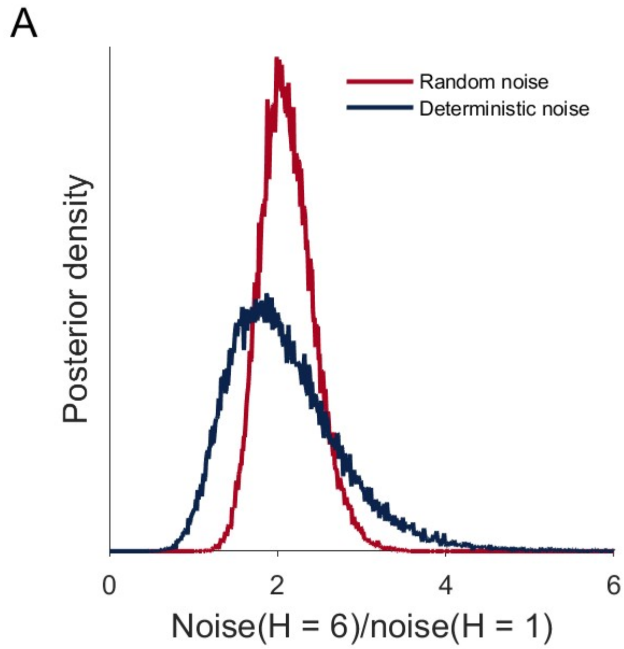


Figure 8: Model based analysis showing the posterior distributions over the ratio of the group-level mean of the standard deviations of random and deterministic noise between horizon 6 and horizon 1 respectivelly. The ratio in the standard deviations of noise between horizon 6 and horizon 1 is similar for random and deterministic noise.

Posterior predictive checks

In addition to fitting the model to behavior, it is also important to check whether the model captures the qualitative patterns of the data (**?**) — specifically how p(high info), p(low mean) and p(inconsistent) change with horizon.

To perform this 'posterior predictive check,' we created a set of simulated data by taking the subject-level parameters from the hierarchical Bayesian fits and having the model play the same sequence of games as seen by the subjects. We then applied the same model-free analysis as described in the previous sections to this simulated data set and compared the model's behavior to that of participants. As shown in Figure 9, the model can account for all qualitative patterns in the data — the increase in p(high info), p(low mean), and p(inconsistent) with horizon, and that p(inconsistent) is in between pure random and pure deterministic noise. The quantitative agreement is almost perfect for p(high info) and for p(inconsistent) in the [1 3] condition, but the model seems to systematically overestimate p(low mean) and p(inconsistent) in [2 2] conditions, although the discrepancy is relatively small (overestimating p(low mean) by 0.054 or

31.37%, and p(inconsistent) by 0.049 or 27.83% in [2 2] condition).

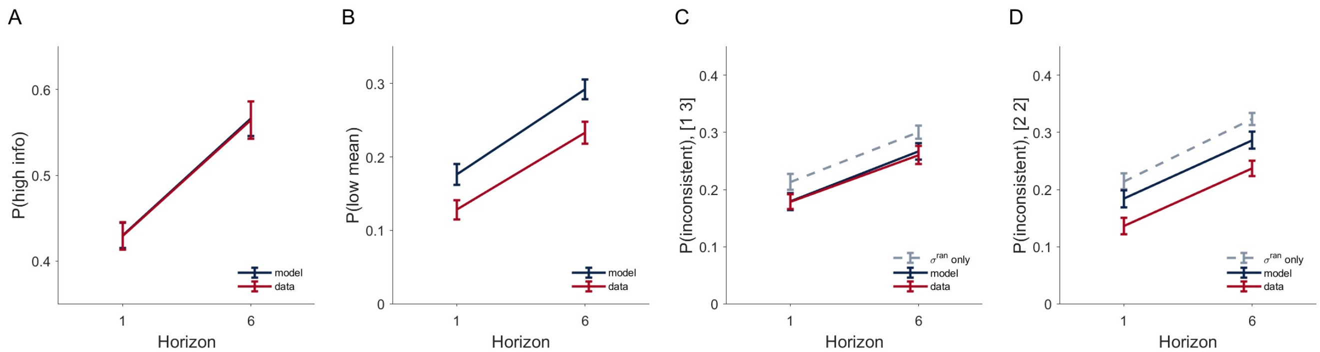


Figure 9: Our model accounts for all qualitative patterns of the data, namely, p(high info) and p(low mean) increase as a function of horizon, p(inconsistent) increases as a function of horizon for both [1 3] and [2 2] conditions and lies between the pure random and pure deterministic noise prediction.

Comparison with alternative models

COULD POTENTIALLY MOVE THIS TO THE SUPPLEMENT

To check whether all aspects of the model were necessary to reproduce the qualitative pattern of findings, we also built and fit five additional versions of the model. These models varied in[siyu] whether deterministic and random noise are present or not and whether either types of noise is dependent on horizon. Specifically, we tested the following 6 models (Note that the $\sigma^{ran}_{horizon}, \sigma^{det}_{horizon}$ model is our original full model).[siyu]

Model	Deterministic noise	Random noise
$\sigma^{ran}_{horizon}, \sigma^{det}_{horizon}$	Horizon dependent	Horizon dependent
$\sigma^{ran}_{horizon}, \sigma^{det}$	Fixed	Horizon dependent
$\sigma^{ran}, \sigma^{det}_{horizon}$	Horizon dependent	Fixed
$\sigma^{ran}, \sigma^{det}$	Fixed	Fixed
$\sigma^{ran}_{horizon}$	Horizon dependent	None
$\sigma^{det}_{horizon}$	None	Horizon dependent

Table 1: Variants of the model.

The posterior distributions over the group-level means of the deterministic and random noise standard deviation σ_{det} and σ_{ran} (when they exist) in these model variants are shown in Supplementary Figure

S9. We again simulated choices using fitted parameters from these models and repeated the model-free analysis on the simulated data. siyu As shown in Supplementary Figure S10, only one of these models, where random noise is horizon dependent but deterministic noise is not, can capture the full qualitative pattern of behavior. However, the quantitative fit to the data is not as good (Supplementary Figure S10).

[siyu 18] (was): responding

Moreover, we examined if our model can indeed qualitatively capture whether deterministic and random noise are present or not and whether either types of noise is dependent on horizon. To test this, we simulated choices from each of the 6 models, and then fit the simulated choices with our original full model. The simulation was repeated 50 times for each model. Indeed, we showed that our model can capture both the existence of random and deterministic noise, and whether each noise changes with horizon condition (Figure 10), with only one exception that our model falsely detected a small fraction of deterministic noise when no deterministic noise was present (Figure 10, S). This phenomenon was also examined and discussed in the section "Model validation" above. siyu

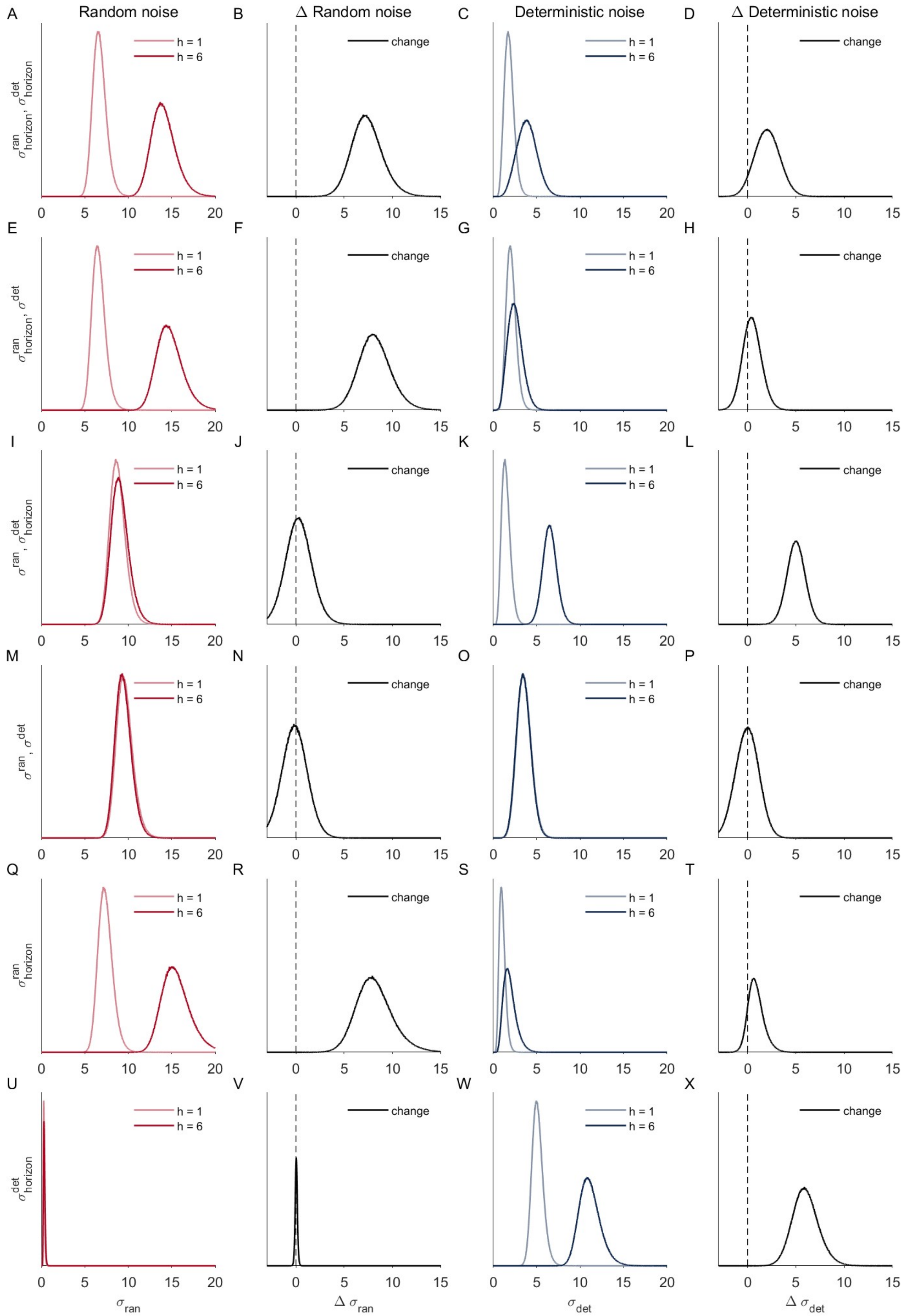


Figure 10: Our model qualitatively captures whether deterministic and random noise are present or not and whether either types of noise is dependent on horizon. A-D. both deterministic and random noise are horizon dependent, E-H. only random noise is horizon dependent, I-L. only deterministic noise is horizon dependent, M-P. neither random nor deterministic noise is horizon dependent, Q-T. only deterministic noise is assumed to be present, U-X. only random noise is assumed to be present.

Discussion

In this paper, we investigated whether random exploration is really random or whether it is driven deterministically by aspects of the stimulus we have previously ignored when measuring 'decision noise.' Using a version of the Horizon Task with repeated games, we found evidence that at least some of the noise in random exploration could be explained by such 'deterministic noise.' In particular, we found that deterministic noise accounted for around 15% of the overall variability in people's behavior.

SUGGEST REFRAMING NEXT PARAGRAPH IN THE FOLLOWING WAY. SIMPLE READ OF RESULTS SUGGESTS RANDOM EXPLORATION IS RANDOM - THIS IS CONSISTENT WITH X. FOLLOWING PARA THEN GIVES LIMITATIONS AND TALKS ABOUT DETERMINISTIC INTERPRETATION. I'VE TRIED TO DO THIS ...

One interpretation for this low level of deterministic noise is that most of the variability in random exploration is truly random. Such a random noise interpretation, would be consistent with recent work showing that variability in perceptual decisions may be driven by imperfections in mental inference **?**. In this view, apparently random behavior is not due to sensory processing or response selection, but to suboptimal computations in the brain. Although suboptimal inference is different from simply adding random noise to neural circuitry(**?**), as long as the suboptimality in neural computation is not a deterministic function of the stimuli, it is a form of random noise in our definition. Indeed, a strong interpretation of this hypothesis would suggest that randomness in explore-exploit behavior is due to imperfect inference about the correct course of action. In the context of the Horizon Task, such computational errors would likely be larger in the long horizon condition as the correct course of action in these cases is much harder to compute CITE DEEP EXPLORATION PAPER.[bob]

While the random noise interpretation is theoretically appealing, our approach, while an improvement on previous methods, is not without limitations. Most important is that our measure of 'random' noise is only an upper bound on the true level of randomness and that, in principle, the random decision noise could be lower. Specifically, in our model, what we labeled random noise was really 'non-stimulus-driven variability.' While this non-stimulus-driven variability could be driven by truly random stochastic processes, it could also be driven by deterministic processing that is unrelated to the stimuli in the task. For example, such deterministic noise could be driven by differences in where people look, or for how long they look, or by whether they were fidgeting or scratching their nose (**?**). In addition to this conceptual limitation in measuring deterministic noise, parameter recovery simulations suggest that our estimation method also slightly underestimates deterministic noise (see Figure 5, Supplementary Figure S6). As a result, from both a conceptual and methodological perspective, it is possible that the remaining 85% of the decision noise that is not stimulus-driven noise, could be deterministic.

Like the random noise account, the deterministic noise account is also in line with previous work in which neural variability can be accounted for by fluctuations in sensory inputs. For example, MT neurons were shown to have a reproducible temporal modulation in response to a fixed random motion stimuli (**?**). In other words, 'irrelevant' features in the stimuli are represented in a reliable way in the brain that could drive downstream choices in a predictable way.

Regardless of whether we interpret the noise as random or deterministic, a key finding in this paper is that both types of noise change with horizon. Such a horizon increase is a hallmark of an exploratory process and suggests that the modulation of deterministic and random processes may underlie random exploration. Moreover, the fact that the horizon change in the two types of noise are proportional to each other (Figure 8) suggests a possible mechanism for random exploration: a reduction in the strength with which reward drives the choice.

To see how a change in reward processing could affect random and deterministic noise, consider the simple decision model we introduced in Equation XXX. In this model, choice is determined by the sign of the difference in utility ΔQ between the two options, where

$$\Delta Q = \Delta R + A \Delta I + b + n_{det} + n_{ran} \tag{3}$$

Now imagine a case where the reward signal is scaled by a factor β. In this case, ΔQ becomes

$$\Delta Q = \beta \Delta R + A \Delta I + b + n_{det} + n_{ran} \tag{4}$$

Because the choice only depends on the sign of ΔQ, scaling ΔQ by a factor of $1/\beta$ will not change the behavior of the model. Thus, if we divide both sides of the above equation by β we get

$$\Delta Q/\beta = \Delta R + A \Delta I/\beta + b/\beta + n_{det}/\beta + n_{ran}/\beta \tag{5}$$

which is equivalent to a scaling of both deterministic and random noise by the same factor $1/\beta$. Thus, one interpretation of our result that both deterministic and random noise change across horizons with the same ratio, is that this reflects a change in reward processing. That is, the reward signal is reduced in the longer horizon condition (smaller β in horizon 6 than horizon 1).

Such a reduction in the strength of reward coding in exploration, is consistent with our recent work using a drift diffusion model (DDM) to model explore-exploit decisions CITE FENG ET AL. In the drift diffusion model, changes in behavioral variability can be driven by changes in the decision threshold (smaller threshold = more noise) or changes in the signal-to-noise ratio with which reward is encoded (lower SNR = more noise). By fitting both choices and response times, we were able to distinguish between these two accounts showing that the majority of the horizon-change in variability was driven by changes in SNR not threshold. However, this model could not determine whether the changes in SNR were driven by signal or noise. By showing that the change in deterministic and random noise have the same ratio, the present work suggests that this SNR change is driven by changes in reward-signal processing, not noise. Of course, to truly see whether changes in signal or noise are driving random exploration will require more direct measurements of neural processing such as with neuroimaging and electrophysiology (CITE TOMOV GERSHMAN FMRI, CITE EBITZ).

Materials and Methods

Participants

80 participants (ages 18-25, 37 male, 43 female) from the University of Arizona undergraduate subject pool participated in the experiment. 14 were excluded on the basis of performance, using the same exclusion criterion as in (**?**). In this exclusion criteria, we measured the accuracy of each participant's choices by calculating the percentage of times that a participant chose the bandit with the higher underlying mean payouts in the last choice of a long horizon game, intuitively people should figure out which bandit has a higher mean payout by the last trial and should have an accuracy measure significantly above 50%, specifically, we computed the likelihood that the measured accuracy can be achieved by making a completely random choice between the two options and excluded participants with a likelihood smaller than 99.999%, in other words, participants who didn't show an accuracy significant above chance with $p < 0.001$ were excluded in the analysis. This left 65 for the main analysis. Note that including the 15 badly performing subjects did not change the main results (Supplementary Figures 1 - 3)

[siyu 19] (deleted): driven

Task

The task was a modified version of the Horizon Task (**?**) (Figure 1). 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. 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 Gaussians were different for the two bandit options, remained the same within a game, but changed with each new game. One of the bandits always had a higher mean than the other. Participants were instructed to maximize the points earned over the entire task. 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.

The number of games participants played depended on how well they performed, which acted as the primary incentive for performing the task. Thus, the better participants performed, the sooner they got to leave the experiment. On average, participants played 153.7 games (minimum = 90 games, maximum = 192 games) and the whole task lasted between 12.37 and 32.15 minutes (mean 22.78 minutes). Participants played an average of 65.3 repeated pairs of games (minimum = 30 repeated pairs, maximum = 79 repeated pairs).

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. 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. Participants were informed that in every game one of the bandits always has a higher mean reward than the other. The order of games was randomized. Mean sizes and order of presentation were counterbalanced.

Each game consisted of 5 or 10 choices. Every game started with a fixation cross, then a bar of boxes appeared indicating the horizon for that game. For the first 4 trials - the instructed trials, we highlight the box on one of the bandits to instruct the participant to choose that option. On these trials, they have to press the corresponding key to reveal the outcome. From the fifth 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 participants 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 300ms. The task was programmed using Psychtoolbox in MATLAB (**??**).

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 instructed 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), Figure 1.

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 deterministic noise from random noise.Because the stimuli are identical in the repeated games, by definition, deterministic noise remains the same in repeated games, whereas random noise can change.

[siyu 20] (was): between components of the noise that are deterministically driven by [...]

Hierarchical Bayesian Model

To model participants' choices on this first free-choice trial, we assume that they make decisions by computing the difference in value ΔQ between the right and left options, choosing right when $\Delta Q > 0$ and left otherwise. Specifically, we write

$$\Delta Q = \Delta R + A\Delta I + b + n_{det} + n_{ran} \tag{6}$$

where, the experimentally controlled variables are $\Delta R = R_{right} - R_{left}$, the difference between the mean of the rewards shown on the forced trials, and ΔI, the difference of 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. The other variables 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_{det} and n_{ran} are deterministic noise and random noise respectively. n_{det} denotes the deterministic noise, which is identical on the repeat versions of each game; and n_{ran} denotes random noise, which is uncorrelated between repeat plays and changes every game.

Each subject's behavior in each horizon condition is described by 4 free parameters (Table 2): the information bonus A, the spatial bias, b, the standard deviation of the deterministic noise, σ_{det}, and the standard deviation of the random noise, σ_{ran}. 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

(Figure 11). 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 horizon condition i and for each participant s, 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, and σ_i^A from an Exponential distribution with parameters 0.1.

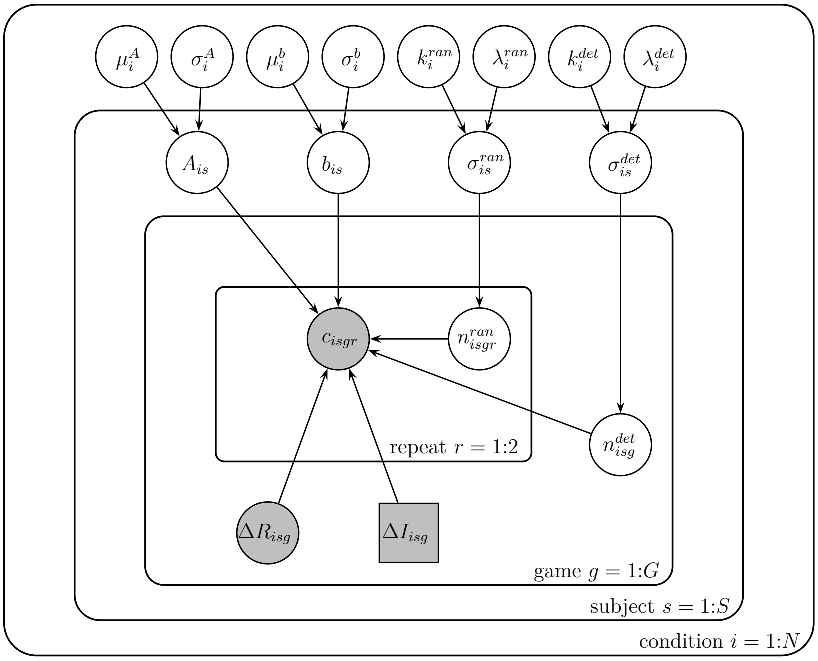
Parameter	Prior	Hyperparameters	Hyperpriors
information bonus, A_{is}	$A_{is} \sim$ Gaussian(μ_i^A, σ_i^A)	$\theta_i^A = (\mu_i^A, \sigma_i^A)$	$\mu_i^A \sim$ Gaussian(0, 100) $\sigma_i^A \sim$ Exponential(0.01)
spatial bias, b_{is}	$b_{is} \sim$ Gaussian(μ_i^b, σ_i^b)	$\theta_i^b = (\mu_i^b, \sigma_i^b)$	$\mu_i^b \sim$ Gaussian(0, 100) $\sigma_i^b \sim$ Exponential(0.01)
deviation of deterministic noise, σ_{isg}^{det}	$\sigma_{is}^{det} \sim$ Gamma(k_i^{det}, λ_i^{det})	$\theta_i^{det} = (k_i^{det}, \lambda_i^{det})$	$k_i^{det} \sim$ Exponential(0.01) $\lambda_i^{det} \sim$ Exponential(10)
deviation of random noise, σ_{isgr}^{ran}	$\sigma_{is}^{ran} \sim$ Gamma(k_i^{ran}, λ_i^{ran})	$\theta_i^{ran} = (k_i^{ran}, \lambda_i^{ran})$	$k_i^{ran} \sim$ Exponential(0.01) $\lambda_i^{ran} \sim$ 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 10 independent Markov chains to generate 50000 samples from the posterior distribution over parameters (5000 samples per chain). Each chain had a burn in period of 5000 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.



Priors

$\mu_i^A \sim \text{Gaussian}(0, 100)$, $\sigma_i^A \sim \text{Exponential}(0.01)$

$\mu_i^b \sim \text{Gaussian}(0, 100)$, $\sigma_i^b \sim \text{Exponential}(0.01)$

$k_i^{ran} \sim \text{Exponential}(0.01)$, $\lambda_i^{ran} \sim \text{Exponential}(10)$

$k_i^{det} \sim \text{Exponential}(0.01)$, $\lambda_i^{det} \sim \text{Exponential}(10)$

Subject specific parameters

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

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

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

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

Deterministic noise for repeated game

$n_{isg}^{det} \sim \text{Logistic}(0, \sigma_{is}^{det})$

Random noise for each game

$n_{isgr}^{ran} \sim \text{Logistic}(0, \sigma_{is}^{ran})$

Observed choices

$\Delta Q_{isgr} \leftarrow \Delta R_{isg} + A_{is}\Delta I_{isg} + b_{is} + n_{isgr}^{ran} + n_{isg}^{det}$

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

Figure 11: Schematic of the hierarchical Bayesian model using notation of **?**

Data and code

Behavioral data as well as MATLAB codes to recreate the main figures from this paper will be made available upon publication.