Barcelona School of Economics

Prediction Competition Final Paper

by

Saleh Fayaz

Professor Gaël Le Mens

Department of Economics and Business at UPF

Professor Mikhail Spektor

Department of Economics and Business at the UPF

March , 2022

Table of Contents

Chapter 1

Exploratory Data Analysis

The objective of this section is to seek **choice predictors** which are dominant between and within different individuals and also tasks. Proceeding in this way we will first visualize decision story of individuals during each task. By that we will get some intuitions about choice predictors which in turn will help us on designing models.

Assumption 1. *Option's value evaluation in each trial will be made in comparison to the another option, and not solely on its own current value.*

What is a choice? Is it choosing one option, or also it means not choosing the another one? If we are thinking in **decision step** we are choosing one vs the another we may also think that in **evaluation step**, in which we update our perceived value of each option, we consider the another one. 1 is also reasonable since:

- **Context Dependent** models were more successful due to large-scale choice predictions. [1]
- The more unstructured the task, more people will simplify it to take a picture of it. So thinking about which one is better needs only one memory place, whereas considering both at the same time demands for two.

Model Element 1. *Choice in each trials is a function of previous differences observed between B and A option(having A as ground).*

Value direction at trial i:

$$VD_i = \text{Value of Option B - Value of Option A}$$

Choice at trial t:

$$Choice_t = F(VD_1, ...VD_{t-1})$$

Which in subsequent chapters we have $Choice_t = 1$ *for option B and 0 otherwise.*

1.1 Picture of Choices

I visualize individuals choice strategy in task 3 as an example 1.1. (the other ones can be seen on *Appendix A*)

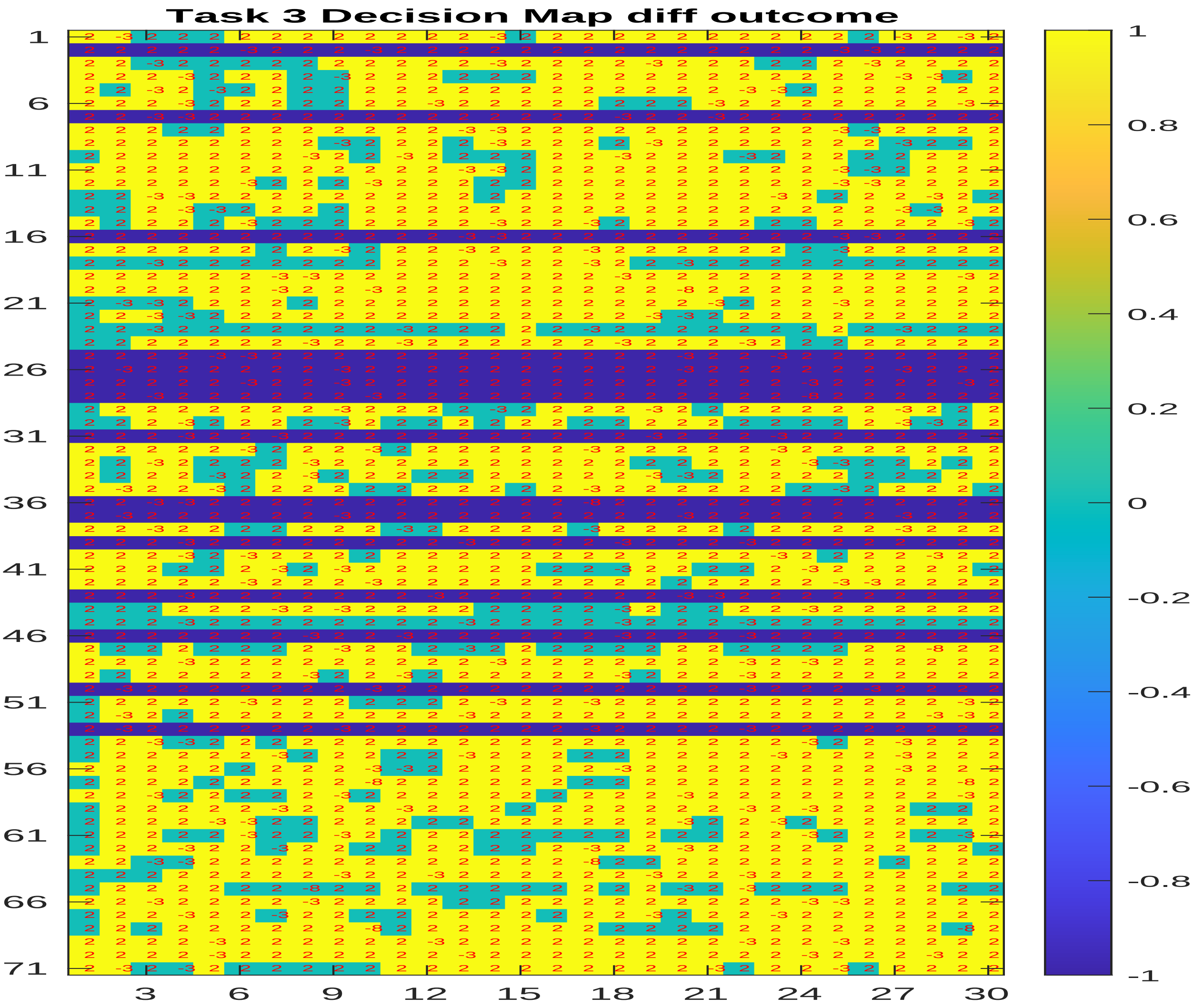

Figure 1.1: Task 3 Decision Map: Each row represents each individual behaviour in this task and each column represents a trial. Color code of value 1 is for individual choosing option B and 0 otherwise. we don't have the choice for -1's. Values in each square represents VD_i seen by individual at trial t.)

As you can see people change their decision suddenly in a simple task as 3 which there is an obvious dominance of an option. In next section I will briefly discuss how people's choice could differ based on current VD.

1.2 Choice Type Distribution

In the literature for the sequential choice task, specifically Bandit problems, there is a substantial evidence for *Win-Stay Lose-Shift Model* [2] in which people stay of shift their choice solely based on win or loose on each trial(Recency dominance). So to understand different behaviours and individuals I define four kind of behaviour as follows:

1. **Persist Based on current Value**: Don't change your decision if current VD agrees with your previous choice.
2. **Ignore current reward**: Don't change your decision even if current VD contradicts with your previous choice.
3. **Change Based on Current Reward**: Change your decision if you contradictory evidence of your previous choice.
4. **Change Based on Current Reward**: Change your decision if you contradictory evidence of your previous choice.

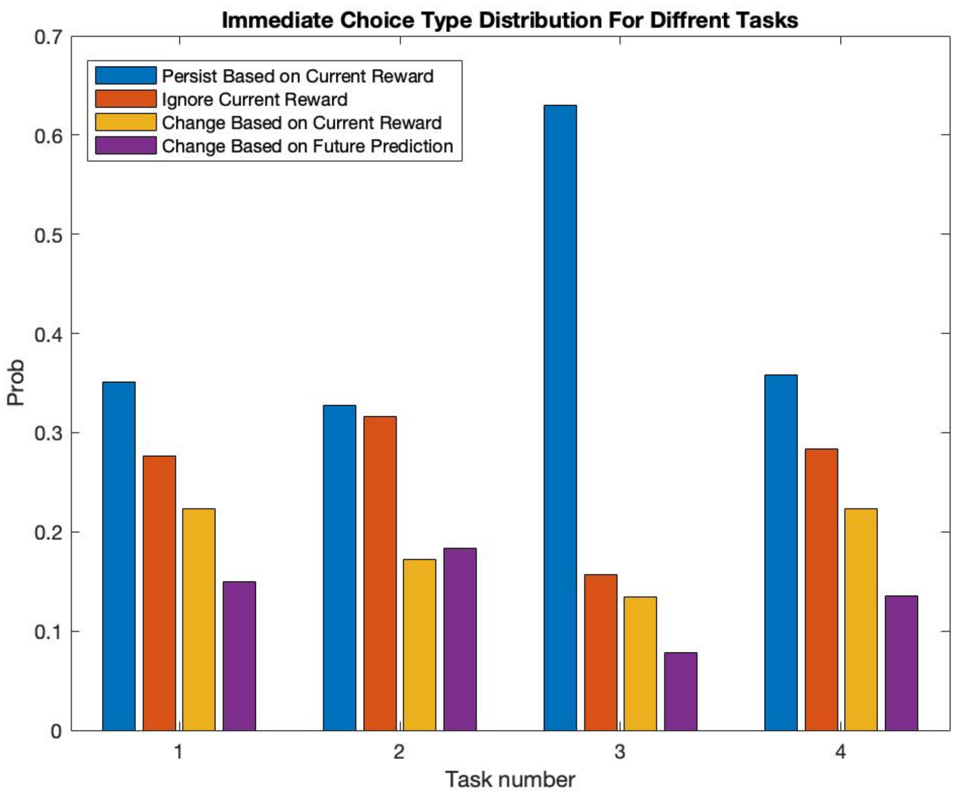

Figure 1.2: Choice type distribution

Figure 1.2 how choice type distribution differs across tasks. Most important point here is that **Task 3 differs substantially from other tasks and task 4, even though has dominant option, is similar to task 1 and 2.**.

Issue 1. ***People will show different characteristics of themselves on each task and this may cause substantial problem if you want to predict a task based on dissimilar one.***

For example If you want to predict task 2 based on task 3 choices then you maybe more biased toward thinking that your individuals are showing more risk-avoidance behaviour, since this is major choice behaviour in task 3.

I would try two resolve Issue 1 in major two ways:

1. Instead of summing up Likelihood, use a weighted version of that in which each choice is weighted based on distribution of task that you want to predict. For example if you want to predict task 2 based on task 3 then give less weight to the choice type 1(which is more dominant in task 3) in likelihood optimization problem.

2. Try to learn an average strategy among individuals on each task which will contribute in individuals choice, separately from his/her performance on other tasks.

It will turn out that the second way on its own is dominant.

1.3 Population Behaviour During Different Trials of Each Task

In subsequent figures I will explore how population behaviour will change during each task. By population I mean here considering all individuals behaviour together. Here I will only describe the figures, and in subsequent chapters will choose main intuitions one could get for designing a model.

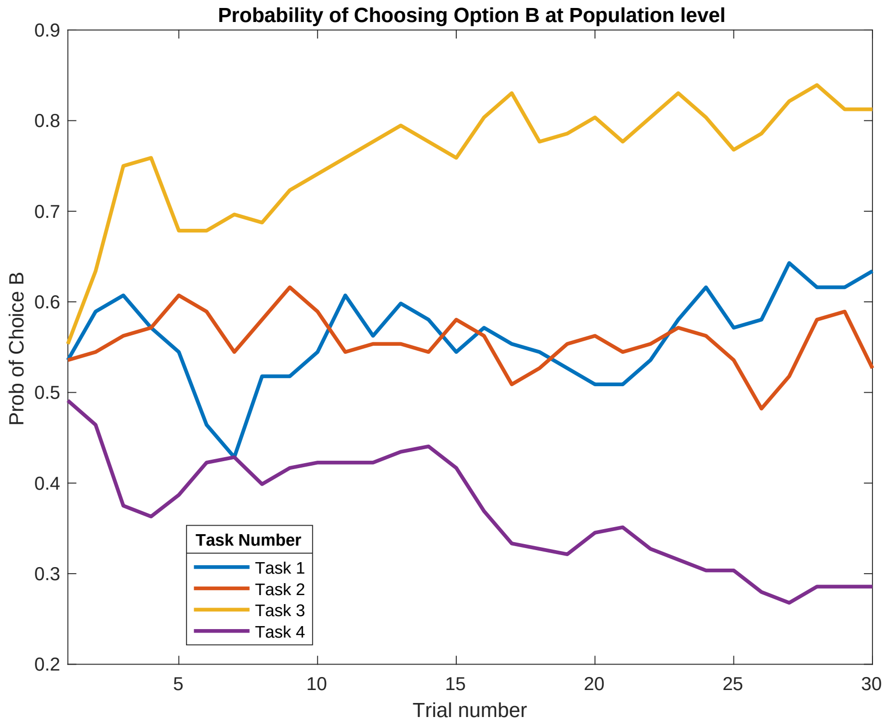

Figure 1.3: Population Behaviour: Probability of choosing Option B varies over time in each task. It seems that people have learnt on average the dominant option in task 3 and 4.

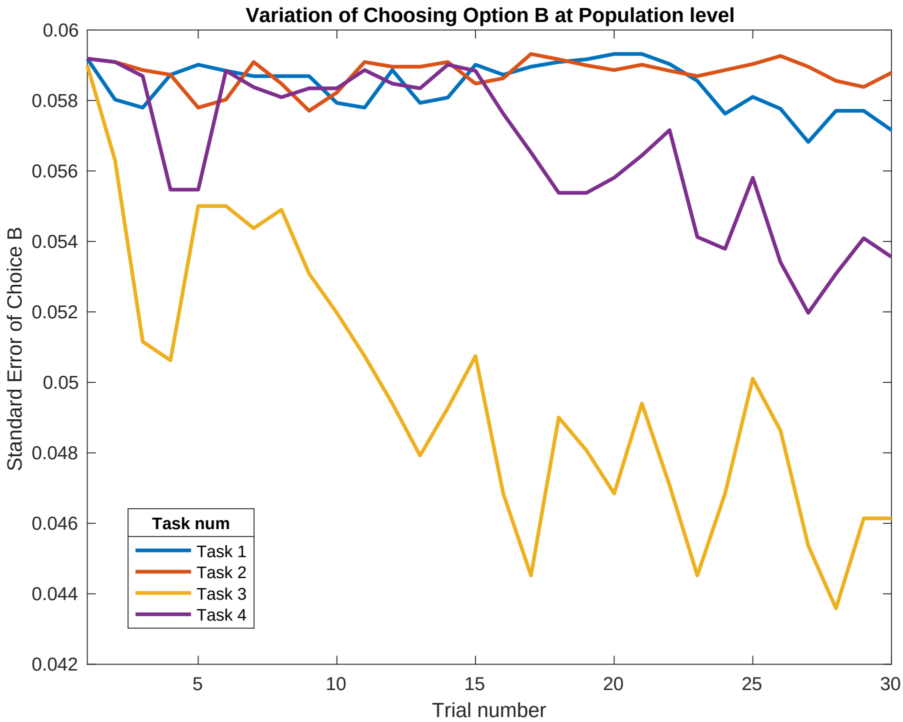

Figure 1.4: Variation across individuals: It seems that task 1 and 2 are showing constant variation in behaviour of individuals where-as in task 3 and task 4 ,which we have dominant option, people ,like each another, rely more in one option than the another.

Another important question could be how **people's choice type** will differ during the task. Are they learning in one period and applying that knowledge on the another? For example if you check choice map of task 3 there is this dominant behaviour that people at the beginning explore more than the near the end 1.1. For accessing this information I will ask what is the probability of changing your choice in each trial. By that I mean: if for example you have chosen option B in previous trial, what is the probability of changing your choice to A in subsequent trial, regardless of immediate reward you had?

We will predict that near the end people have formed a theory about how each task is and react more based on their theories than the immediate reward. By this we may get a measure for learning in each task. 1.5

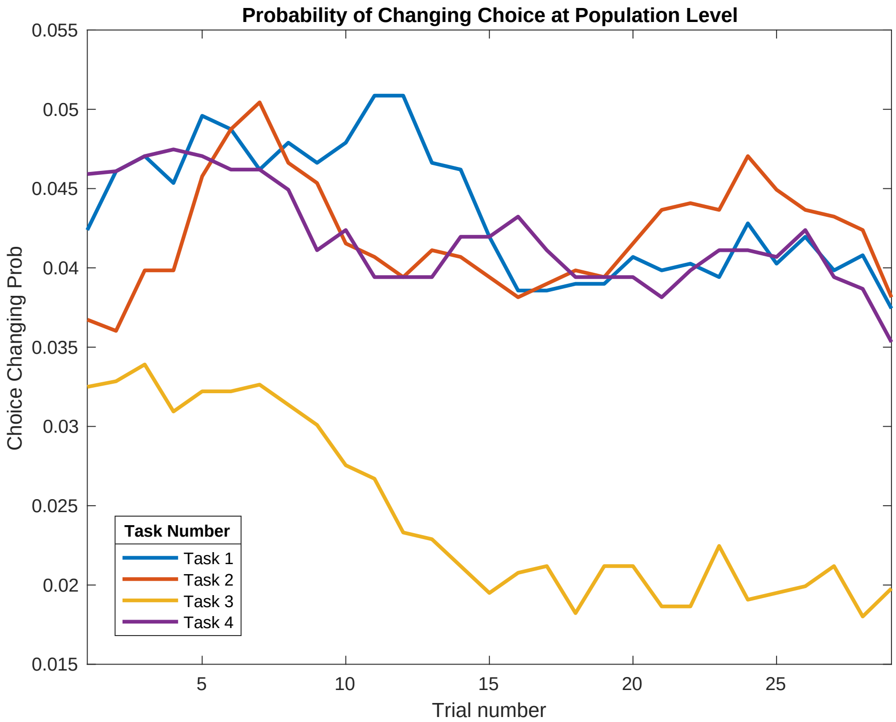

Figure 1.5: Choice Change probability during trial: Most important thing to note here is that people's probability of changing they behaviour in task 4 is more similar to the task 1 and 2 than 3, although we have dominant option!

As I will formulate this in next chapter more formally, this is apparent in choice maps that people show some kind of inertia in their behaviour. in next figure we will explore inertia length assuming that first in 15 trials people will explore more and the next period they will more exploit their knowledge. (**Pretty much interesting!** If you look at population behaviour for task 3 and 4 in the figure 1.4 and figure 1.3 you will find clear evidence for the importance of this number, 15. As it seems after finishing half of the task peoples variation in choice will change suddenly.)

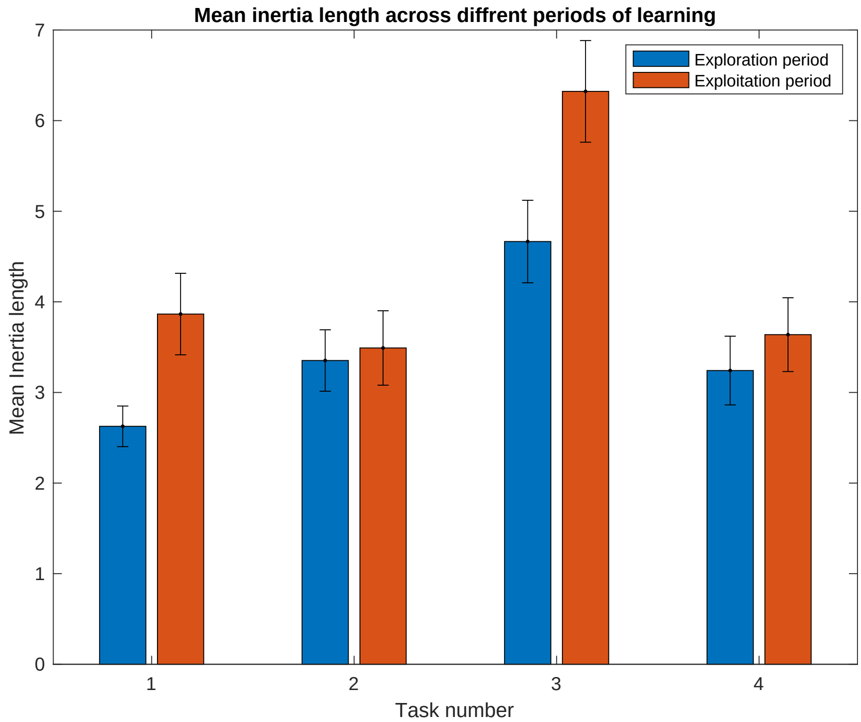

Figure 1.6: Mean inertia length during each period: you can see inertial length differs mostly for the task 1 and 3 but remains same for the task 2 and 4(*very interesting!*)

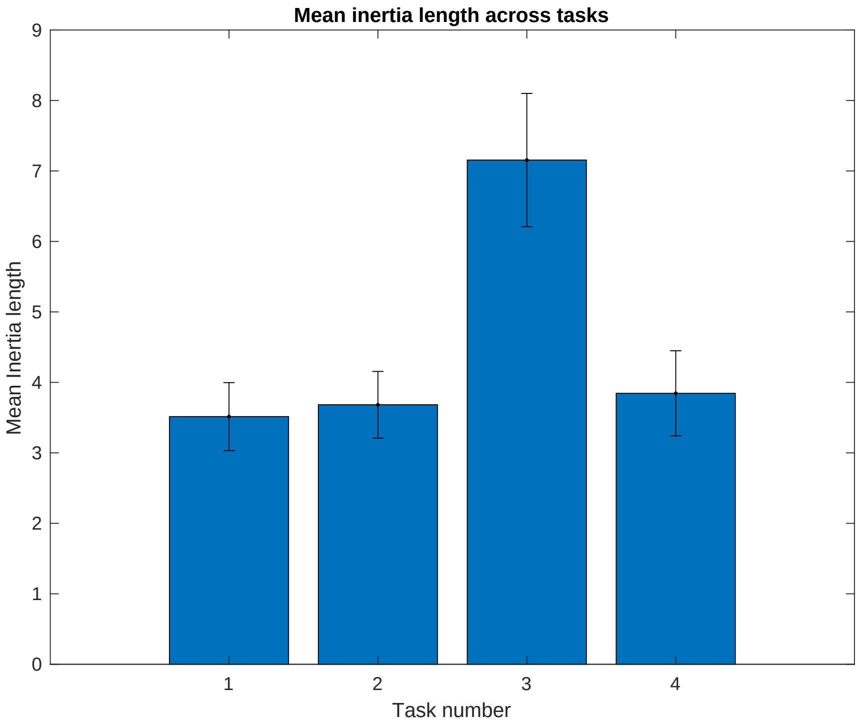

Figure 1.7: Mean inertia length across tasks: same as 1.6

Note!: We have to be cautious interpreting the difference of behaviour across tasks, since structure of each task in itself was different from the first place which in return could be the reason for these differences of behaviour across tasks. For example inertia length is the most for task 3, which could only be result of task structure in which more positive reward comes each after another in this task, whereas in another ones its not the case. But for my point to get only some intuition this could be important that people are more prone to tremble in task 4 in acting by a formed theory to maximize expectation. In regard I should put on my model variables for things other than accumulating rewards or expectation maximization.

Chapter 2
Model design

2.1 Agent Model Structure

I would first differentiate between two step in decision making in order to incorporate each intuition in related step:(Accumulated value difference is AVD)

1. **Evaluation Step**: How Agent accumulate VD_i to get AVD on each trial
2. **Decision Step**: How Agent makes decision on each trial based on AVD and other possible factors

For **Decision Step**:

1. **Issue 2.** *Agent in the baseline model choose between different options by choosing best one, regardless of their absolute differences. This means in place which we have to be cautious about making prediction we are not.*

   Resolving this issue I would add **Soft-max maximization procedure** in which probability of choosing option B relies on value difference of two option.

   Model Element 2.

   $$Prob[choice = B|\text{Reward accumulation model}] = \frac{1}{1 + exp(-\theta * AVD)}$$

   parameter θ here controls for the degree which an agent is making his decision based on expected maximization. (for θ equals to infinity resulting in usual max)

2. **Is it a gamble over aggregated value or next value?**

   Going with our choice type distribution, in which large number of choices has been made in contrast to immediate reward(type 3 and 4), we should think about factors which may make an agent to act like the **opposite** of the reward maximization procedure, since those are the places that we most probably make wrong decision. So I tried to model this probability of acting in the opposite way in intuition part. Name it here as $Prob[ActbyPrediction]$, As I think people will not consider immediate reward only if they **predict** next reward will be the opposite of the current one.

After estimating this probability previous formula to the next one, which is average between:

$$Prob[choice = B] =$$

$$Prob[choice = B|\text{Reward accumulation model}] * (1 - Prob[\text{Act by Prediction}])$$
$$+Prob[choice = A|\text{Reward accumulation model}] * Prob[\text{Act by Prediction}]$$

2.2 Main Intuitions

I will list basic intuitions I had used for my model based on previous chapter and add one element to the model by each :

Intuition 1. *People over or under perceive differences between options differently.(this is equivalent in thinking about degree of risk aversion)*

Model Element 3. *In each step difference between two options will be accumulated based on next formula as perceived differences:*

$$\textit{Perceived difference at trial } i = sign(VD_i) * abs(VD_i)^{risk-seek} \tag{2.1}$$

the more the parameter risk-seek is the larger the perceived difference between options would be for larger values of VD_i .

Intuition 2. *People will learn more when the outcomes are novel.*

Model Element 4. *Alpha parameter in the baseline will be adjusted based on the novelty of the current outcome. We capture novelty like this: How much **sign** of current difference between options is novel. For each person we have different evaluation of novelty so a parameter, **novelty reaction**, will capture this in following formula: Alpha adjusted 1 =*

$$Alpha * [2 * [\frac{1}{1 + exp(-1 * dist2lastseen)} - 0.5]]^{novelty-reaction}$$

which *dist2lastseen* measures in what distance in the past the current sign has been repeated. The more distanced it is, it means current sign is more novel.

Intuition 3. *People will learn during the trial!*

Model Element 5. *Alpha parameter in the baseline will be adjusted based on the trial number as follows to capture how much people will the across trials of a task. We will assume as trial increases people will learn less from the immediate reward, so alpha parameter should decrease with it. Alpha adjusted = Alpha adjusted 1*

$$*exp(-learn_{EM} * (trial_{num}/30))$$

which $learn_{EM}$ measures how much learning happens across trials.

In next two intuitions I will discuss how to determine Prob[Act by Prediction].

Intuition 4. *The more confident A decision is the, less probable you will act by prediction.*

Intuition 5. *People will differ in degree to which they will act by prediction.*

I will include both intuitions in subsequent model element:

Model Element 6.

$$Prob[\text{Act by Prediction}] = (Prob[\text{Opposite reward}]*$$
$$\text{Entropy of previous choice probabilities})^{acpred_{sen}}$$

which parameter $acpred_{sen}$ measures how much concave this function is, which in regard shows how it changes near 0 and 1 values.

Also Entropy of previous choice probabilities here means that if our model predict values surely for the previous choice, confident choice, then probability of acting by prediction should goes to zero. (As we know Entropy will be maximized when p = 0.5 and minimized for p=0 or p= 1)

2.2.1 *How to determine Prob[Opposite reward]?*

An agent will see different outcomes in different trials. But If you see a pattern of outcome, especially at the beginning of task, like this that every 3 trial(sth like inertia length in task 1,2,4) reward will change, then you may act like the opposite of the way reward is signalling. This is not only about the pattern, also if you see many positive values of difference for one option it is more probable that you will go for the another one. And it becomes even worse if he/she cant find the reward pattern change he predicts, since he will predict even more surely that next one will be the in the opposite of direction of current one. (this will happen less in task three which inertia of rewards happens in the task, but in fact this is the difference with task 4 also that inertia in outcome is less than even task 2, which direction of difference between outcomes will differ more often. This is the reason why task 4 behaviours are not like the third task.)

So in order to capture this for each trial I will ask how much is it probable that next outcome changes from the current one. for example If I see that, at mean, every 3 4 trial outcome will reverse, then its more probable for me to deviate after seeing 3 4 same direction of outcome toward one option.

So I would say, if mean distance of two previous time that reward was in the opposite direction of the current one = n , and I am seeing for n trial same outcome till the last time it was the opposite, I will put Prob[Opposite reward] = 1 for the next trial! To generalize more, I will put a Gaussian distribution around nth trial, by which I mean, if I am in ith trial with same outcome, I will put

$$Prob[\text{Opposite reward}] = Gaussian(i, n, sigma)$$

which I put sigma here equal to 2. So we are capturing pattern recognition of individuals in this way that around nth trial it is more probable to deviate from current reward.

Chapter 3

Conclusion

3.1 Discussion

3.1.1 *Other approaches used for first two submission*

With the same intuitions as I have said before, I have designed multiple Decision tree's which for each trials tries to use same features as intuitions said. But the results was much worse than the Reinforcement learning Agent I have used for the last two submission, so I only discussed the best one here

3.1.2 *What was the Best model?*

I have discussed a model with its variables, but the question is how much is it successful in prediction of individuals behaviour. Regarding the 1 predicting individuals behaviour in tasks which task 3 is in training set(or not) will make problem. To solve this problem at individual level we can use weighted optimization approach I had discussed there. (Many of the parameters which were not in the core intuitions has been removed from the model to prevent over fitting problem.)

After having prediction for each individual, we may also use second approach to train an agent for each task (irrespective of which person it is) and then try to predict on desired task.

Then the final probabilities would be weighted average of individual and task level prediction as follows:

$$Prob[choice = B] = Prob[choice = B|indiv] * \beta + Prob[choice = B|task] * (1 - \beta)$$

this parameter beta could varies from zero to one.

In order to find best beta I had used prediction data-set of 11 individuals which we had complete observation for them. So in order to access accuracy of model for each beta I predict each task of each individual based on the other three and then take average over all persons and tasks as mean deviance.

I found $\beta = 0$ will result to best estimation! which basically means best estimation for the model happens when I predict irrespective of persons behaviour in other tasks and predict choices solely based on average behaviour in each task.

Appendix A

Supplementary figures

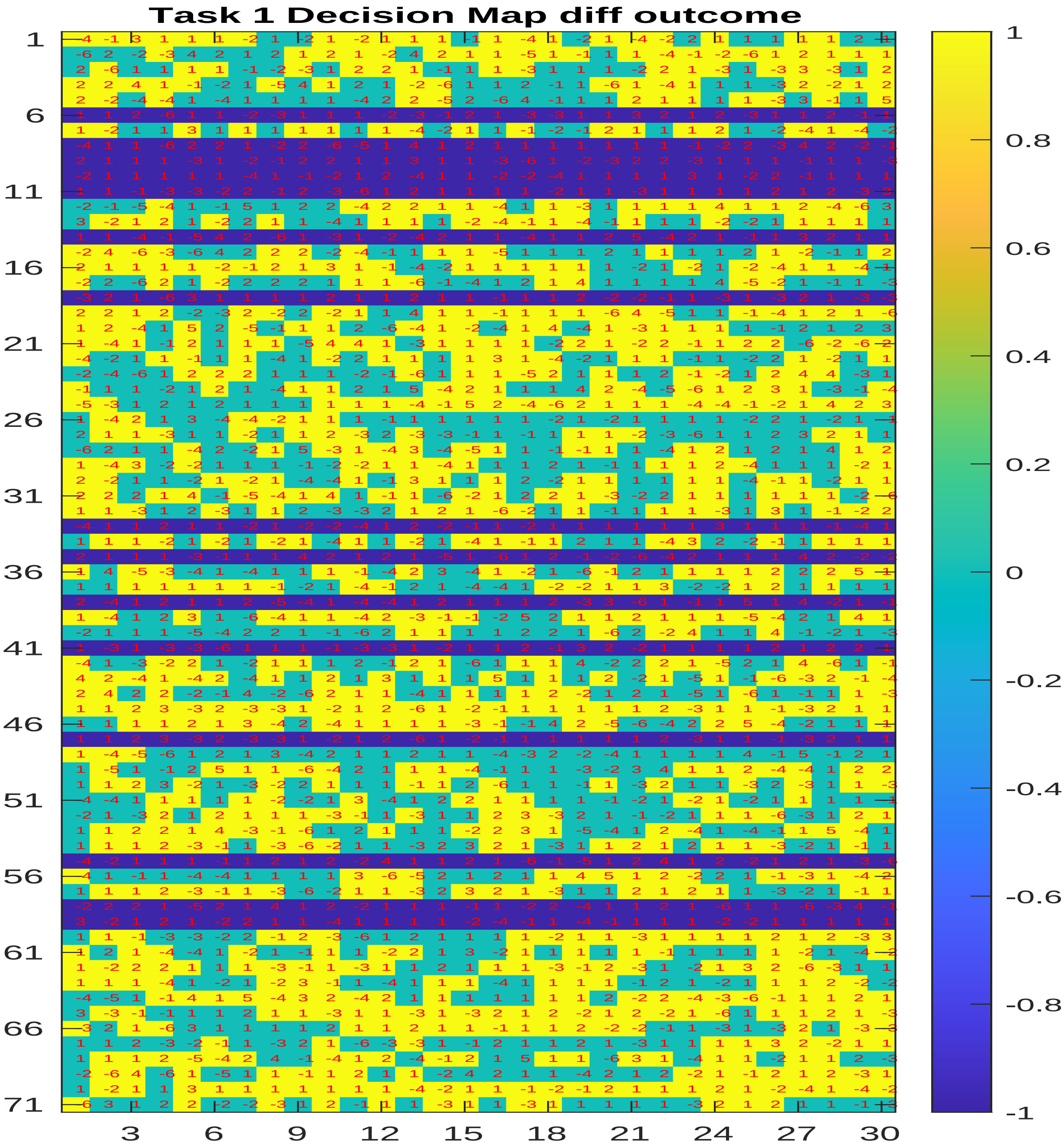

Figure A.1: Task 1 Decision Map

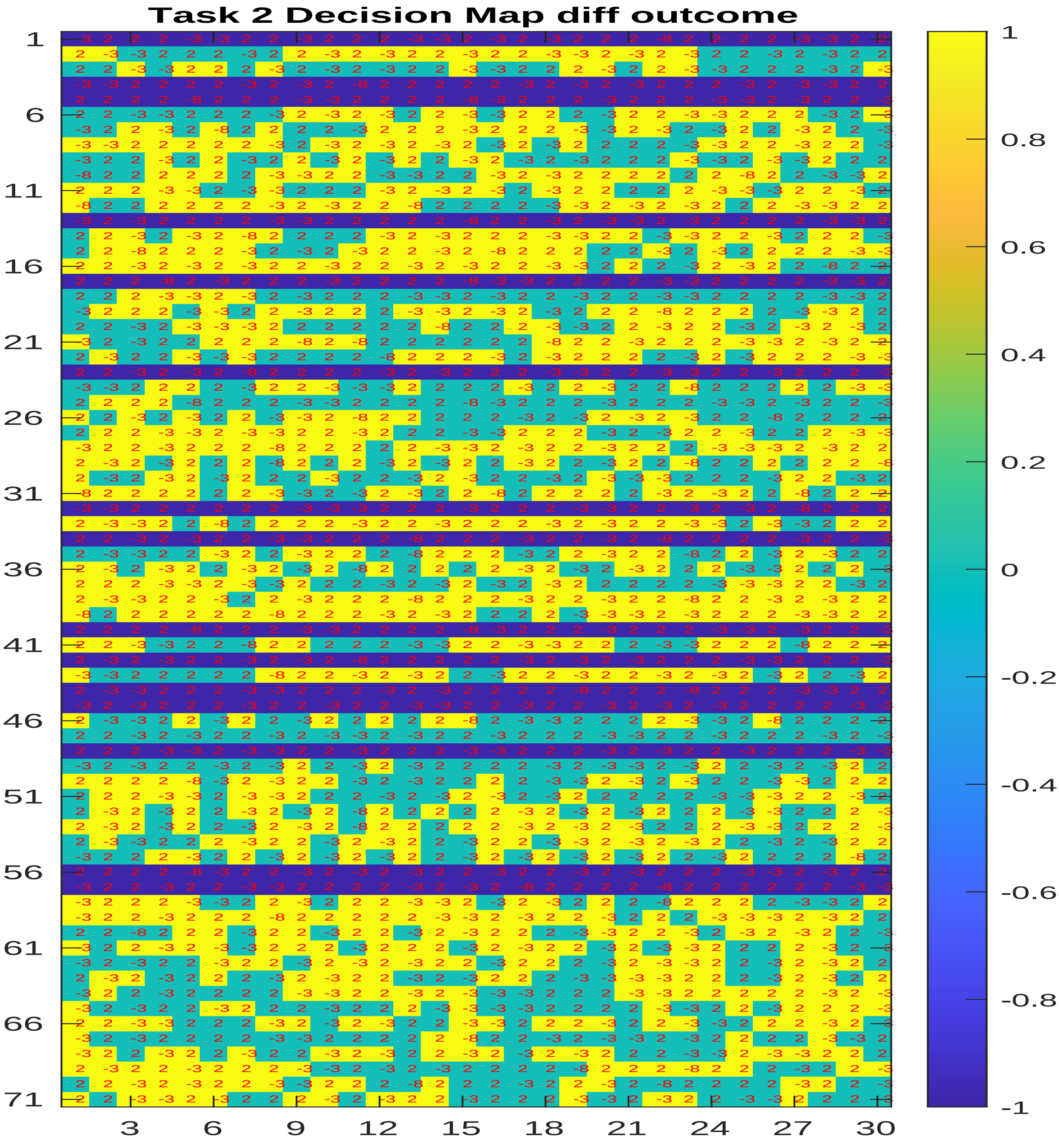

Figure A.2: Task 2 Decision Map

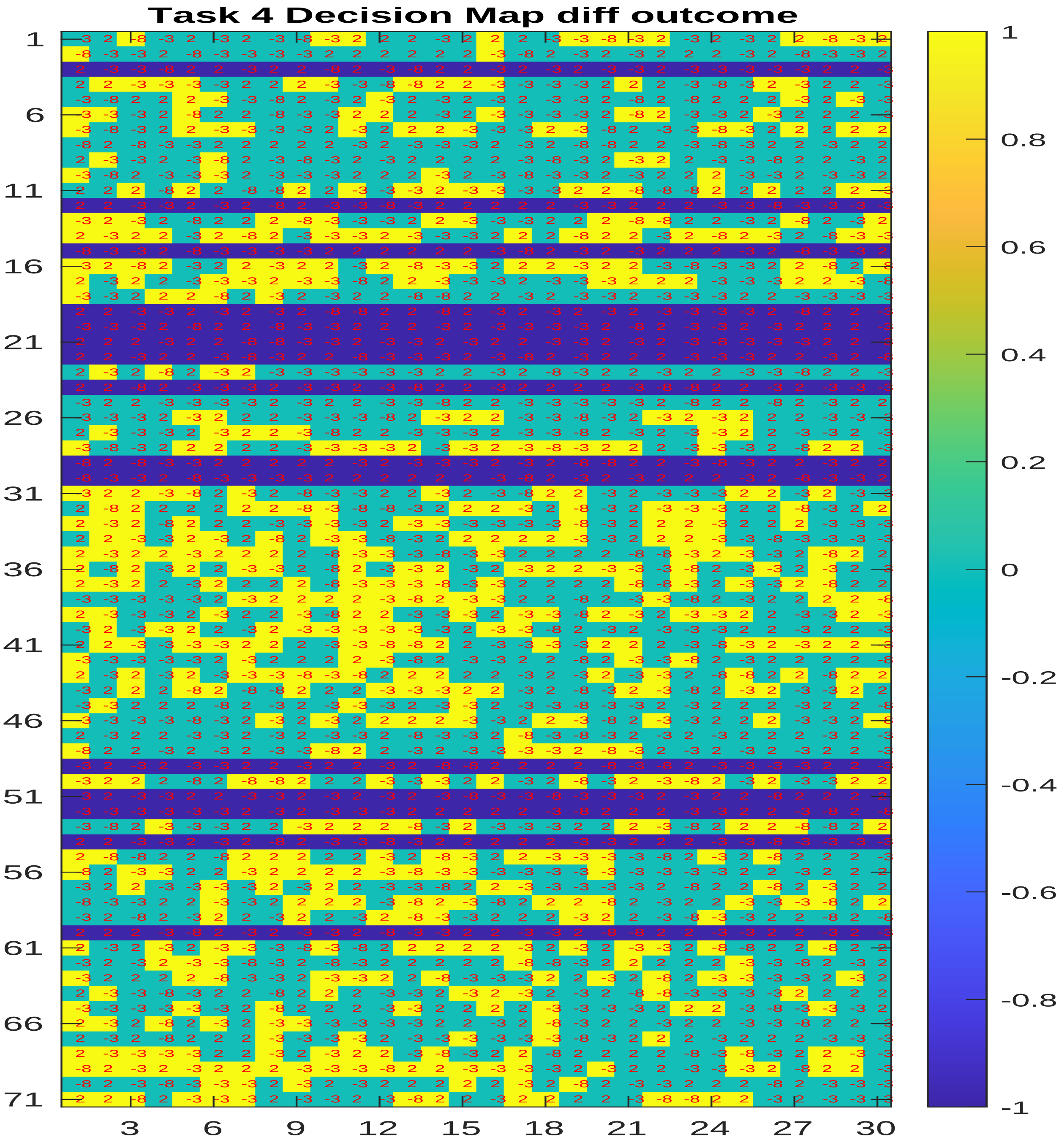

Figure A.3: Task 4 Decision Map

Bibliography

[1] J. C. Peterson, D. D. Bourgin, M. Agrawal, D. Reichman, and T. L. Griffiths, "Using large-scale experiments and machine learning to discover theories of human decision-making," *Science*, vol. 372, no. 6547, pp. 1209–1214, 2021.

[2] M. Guan, *A cognitive modeling analysis of risk in sequential choice tasks*. University of California, Irvine, 2019.