BSE662A - Report

Lab Rats

Arth Banka, 200191
Rashmi G R, 200772
Saloni Das, 200845
Pradumna Awasthi, 200693
Pratyusha Chakraborty, 200718
Lige Nyodu, 22118272

1 Introduction

Decision-making is an essential aspect of everyday life that involves choosing between two options: exploitation, which involves making a decision based on available information, or exploration, which consists of seeking and gathering more information before making a decision. This is also a central problem in reinforcement learning, and understanding the human thought process can help us arrive at efficient solutions.

The replicated experiment involves bandit arms, which can be visualized as slot machines with their own reward distributions. Participants in the experiment must choose which of the four options will result in the highest cumulative reward based on an unfamiliar reward distribution. Wu et al. explore the effects of time pressure on various forms of exploration using a variety of **4-arm (4 payouts) bandit tasks** with varying uncertainty and reward expectations under both time-constrained and unlimited time circumstances. The four payoff conditions were taken from 4 distributions corresponding to IGT, high variance distributions, low variance distributions and the equal means distribution. There was a **time-limited condition of 400 ms and an unlimited time condition**. This also provides a better understanding of regret as a function of reward, and modelling these characteristics can assist in determining estimates of the action value.

One of the key motivators behind our replication was to study the effect of the varied degrees of time pressure people experience when making decisions, referred to as **stochastic time pressure**. One of the main implications of stochastic time pressure is that it might lead to a greater dependence on heuristic and intuitive rather than systematic and analytical decision-making procedures. This is because of the limited cognitive resources available to engage more analytically under a lot of time pressure. Thus participants might act more recklessly and make less accurate decisions. But moderate time pressure might be good for you, too, because it improves your ability to think clearly and make decisions. According to research, **medium time pressure can enhance motivation, focus, and attention to detail, resulting in more precise and effective decision-making**. It is crucial to comprehend how stochastic time constraints affect human decision-making in various contexts, such as banking, healthcare, and aviation, where prompt and accurate decisions can have substantial repercussions. Thus **we have incorporated the 800 ms time-limited condition** in our extension to study how decision-making is affected in this intermediate case. This enables us to explore how time pressure affects the learning rate and how intuition and systematic thinking combine to make choices.

The situational environment can strongly impact the exploration vs exploitation conundrum. The ratio of exploration to exploitation may change based on the circumstances, including risk tolerance, the availability of resources, and the amount of time available. People may be less prone to investigate and take more cautious actions in high-risk situations, whereas they may be more likely to do so in low-risk circumstances. Therefore, a grasp of the situational context is essential for making optimal decisions. It allows individuals to adapt their decision-making strategies to the specific situation and optimize the balance between exploring new options and exploiting known ones. **Our experiment**

design also facilitates studying how situational context affects learning and choice patterns. We have used the condition of a **medical surgery** to simulate the bandit task, with the options presented as surgical tools and the reward presenting itself as health points.

Finally, we have also added **performance bonuses as an additional reinforcement mechanism**, serving both as a motivator for participants to make optimal choices in the trials and as an assessment of their performance in a particular round which the participant can take as feedback to improve their performances.

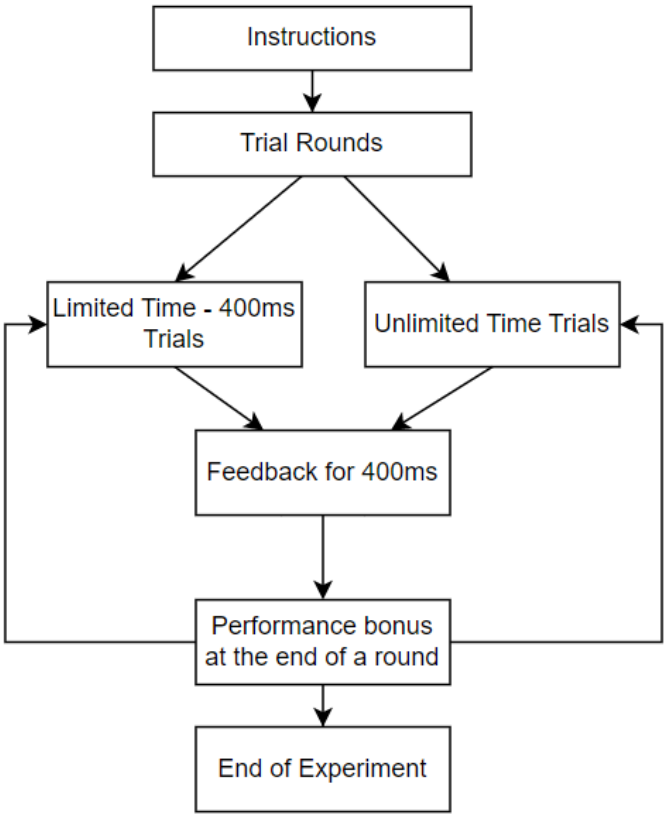


Figure 1: Workflow pipeline for the replication experiment

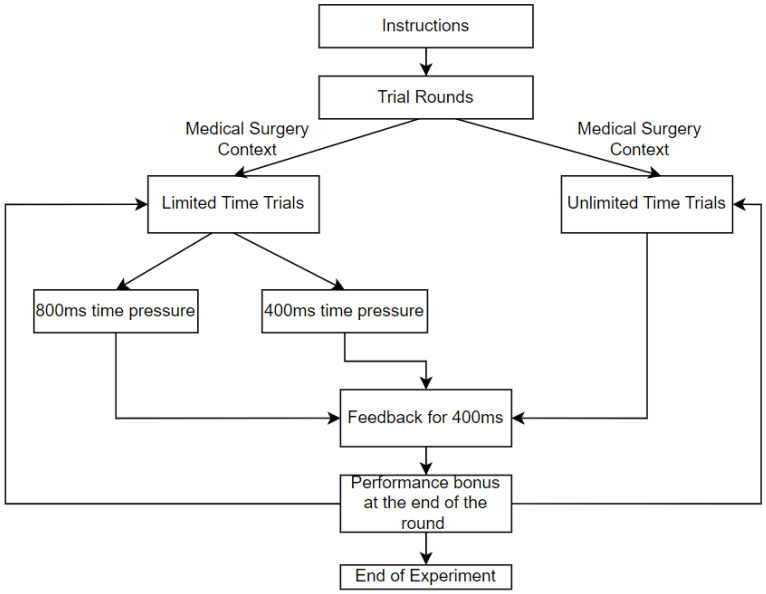


Figure 2: Workflow pipeline for the extension experiment

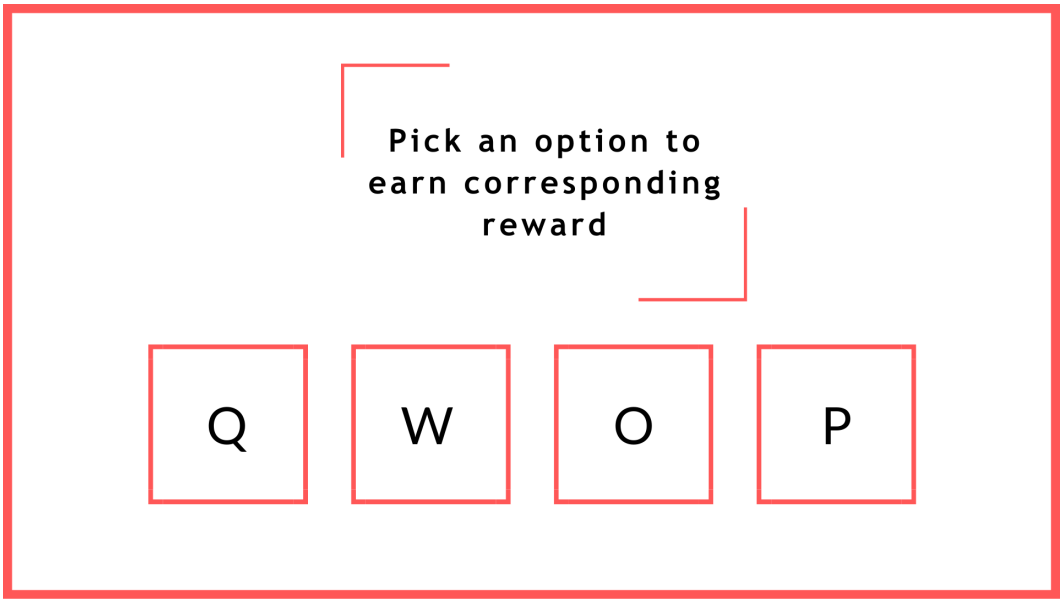


Figure 3: Image denoting the user interface for a participant in the replication experiment

Figure 4: Image denoting how we have added situational context to the extension experiment

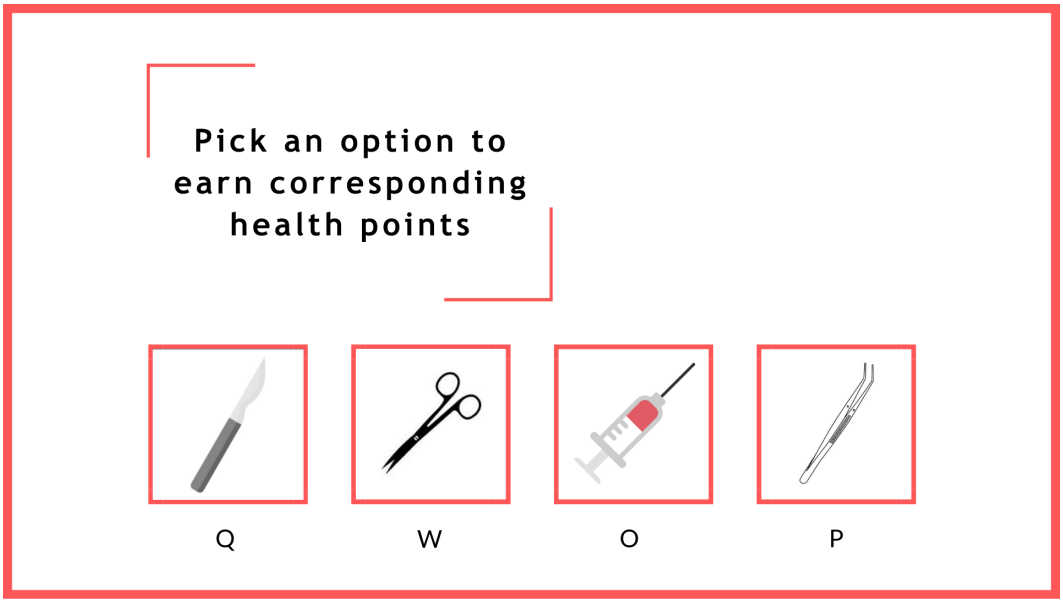


Figure 5: Image denoting how we have presented the options as tools for a surgery to add situational context to the extension experiment

2 Methods

Some main features of the Data Collection Process were :

- A total of 7 common participants were made to play the Replication and Extension Games. This was specifically done to allow for a comparative analysis of the two games. The rest of the 8 participants were made to either play the replication or the extension game but not both.

- It was ensured that the players paid attention to the game strategy and did not randomly press keys to obtain results. This was done mainly because the number of data sets needed to arrive at conclusive outcomes is large but the number and size of available data sets was small.

PLOTS AND ANALYSIS:

The analysis is plot based and the types of Plots used are:

ANALYSIS PLOTS:

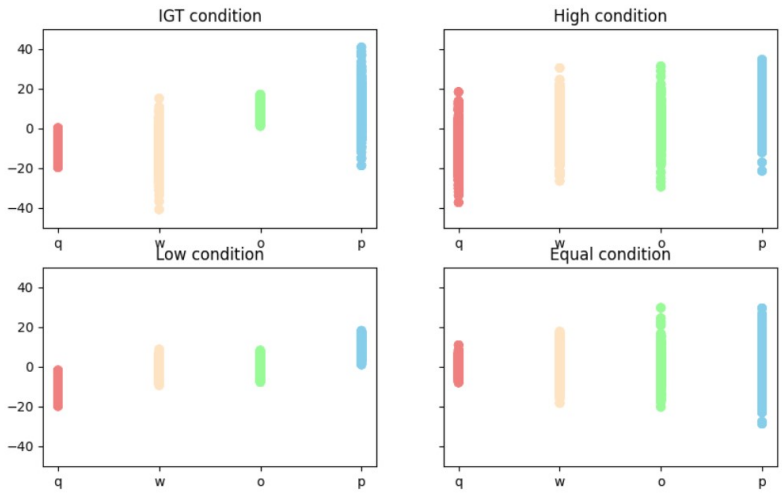
- Payoff Condition Plot: This is a Scatter plot depicting the mean and variance conditions for the four payoff scenarios.
- Learning Curve: This is a curve depicting the average participant learning performance over multiple payoff conditions. The standard deviation is also shown as a lighter ribbon around the main mean curve.
- Choice Entropy Plot: This is a Scatter plot depicting the randomness in participant choices.
- Repeat Choice Scatter Plot: This is a Scatter plot depicting the choice repetition. Each dot shows a participant-wise repeat response probability based on the previous response.
- Arm Frequency Plot: This is a bar plot that shows the random chance of choosing an arm subtracted from the normalized proportion of participant choices over the four payoff conditions.

COMPARATIVE PLOTS:

- Reaction Time Comparison: This plot maps the changes in the reaction times between the 800ms condition and the 400ms condition for the seven common participants.
- Repeat Clicks Comparison: This plot maps the repeat choice trends between the 800ms condition and the 400ms condition for the seven common participants.
- Entropy Comparison Plots: This plot maps the randomness in choice trends between the 800ms condition and the 400ms condition for the seven common participants.

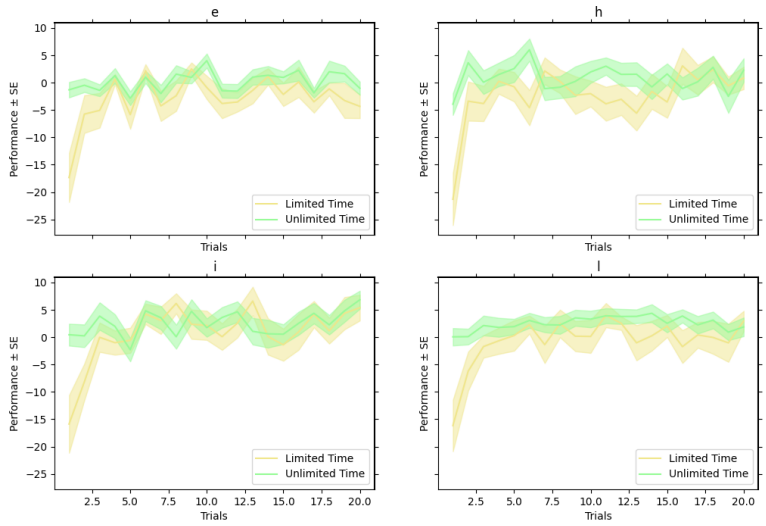
3 Results

3.1 Replication Analysis

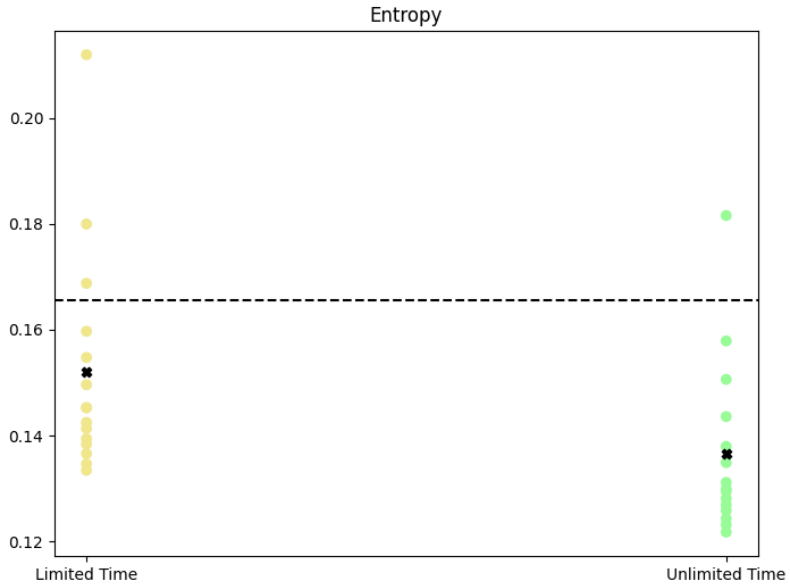


This plot shows the mapping of rewards to the keys q, w, o and p in the various payoff conditions as obtained from the participants' data. The mean and variance of the reward distributions are indicated in the table below.

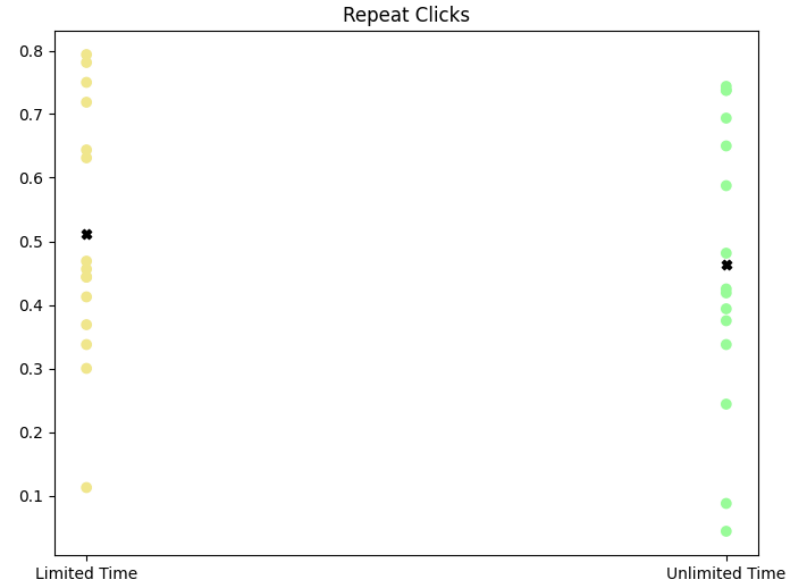
Payoff conds	Means (μ)	Variances (σ^2)
IGT	$[-10, -10, 10, 10]$	[10, 100, 10, 100]
Low var	$[-10, -\frac{1}{3}, \frac{1}{3}, 10]$	[10, 10, 10, 10]
High var	$[-10, -\frac{1}{3}, \frac{1}{3}, 10]$	[100, 100, 100, 100]
Equal means	[0, 0, 0, 0]	[10, 40, 70, 100]



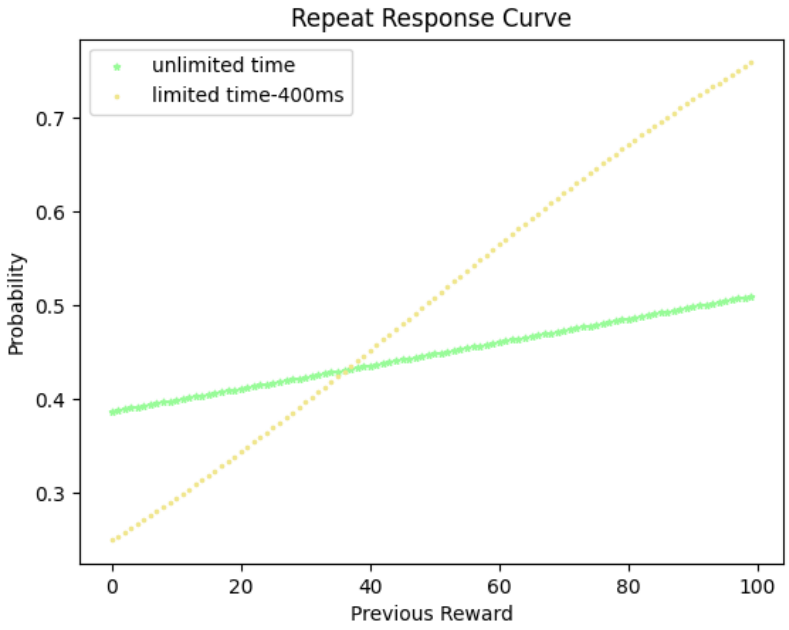
The above plots for the average rewards for each of the payoff conditions under different time conditions show us that the average rewards earned by the participants were in general higher in the case of unlimited time as compared to limited time. However, we did not observe any significant differences when comparing the plots for different payoff conditions.



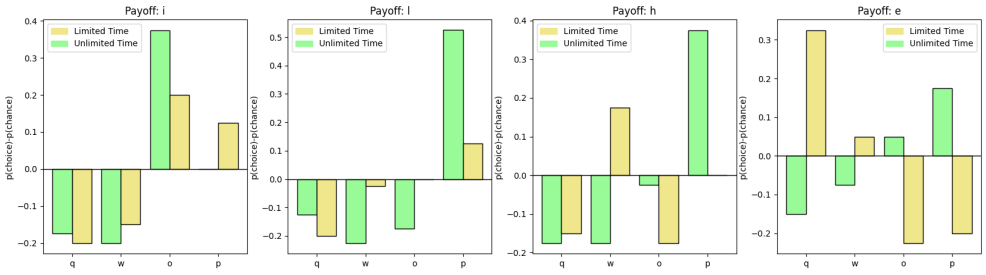
The graph for the choice entropy showed that people made higher entropy or more diverse choices under limited time conditions. This would mean they explored more under time restrictions.



At the same time, it was not consistent when the results were plotted for the repeat clicks as they showed that repeat clicks were also higher in case of limited time conditions which meant that the participants didn't really explore due to risk aversion and went with the same choices.

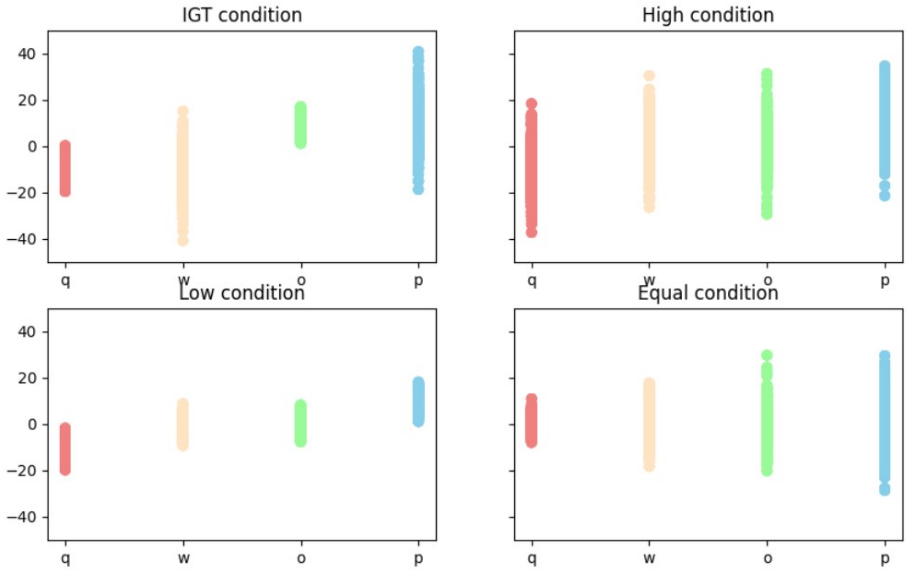


When plotting the repeat choices as a function of the previous reward, we found that people were more likely to repeat their choices if the previous rewards were higher in the case of limited time conditions.

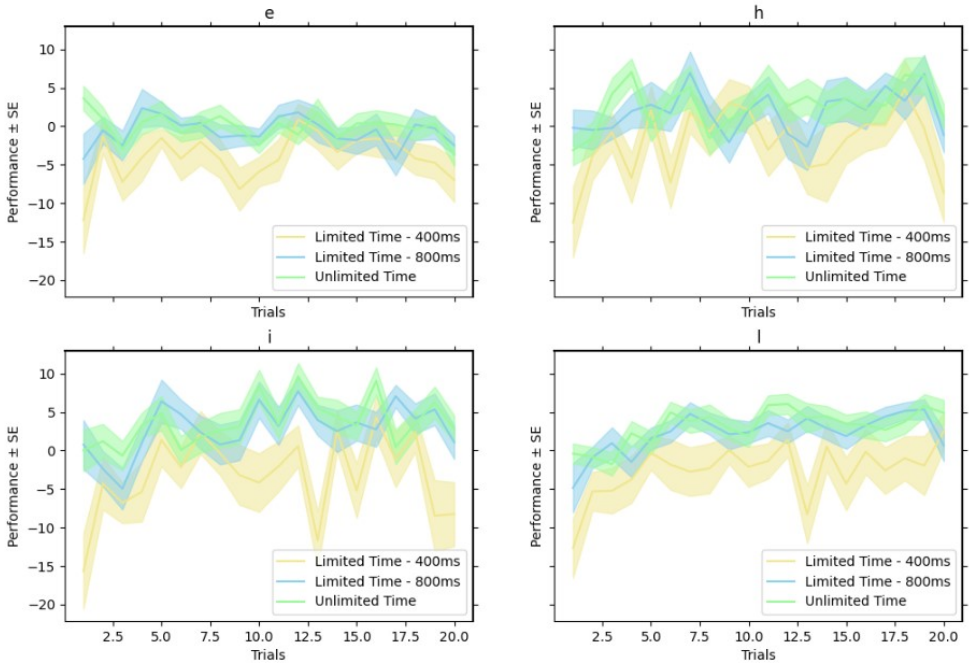


Lastly, we plotted the aggregate choice frequency for each of the options under different payoff conditions. We found that option O was the most frequently chosen option in IGT condition in both the time conditions as option O is a high reward and low variance option. In equal means condition, it is observed that option Q, a low variance option, is chosen frequently in limited time conditions indicating that our participants were more risk averse under time pressure. However, in umilited time, they were more inclined to option P which has high variance or uncertainty, indicating that when there is no time pressure, people were more risk seeking and hence, chose more uncertain options.

3.2 Extension Analysis

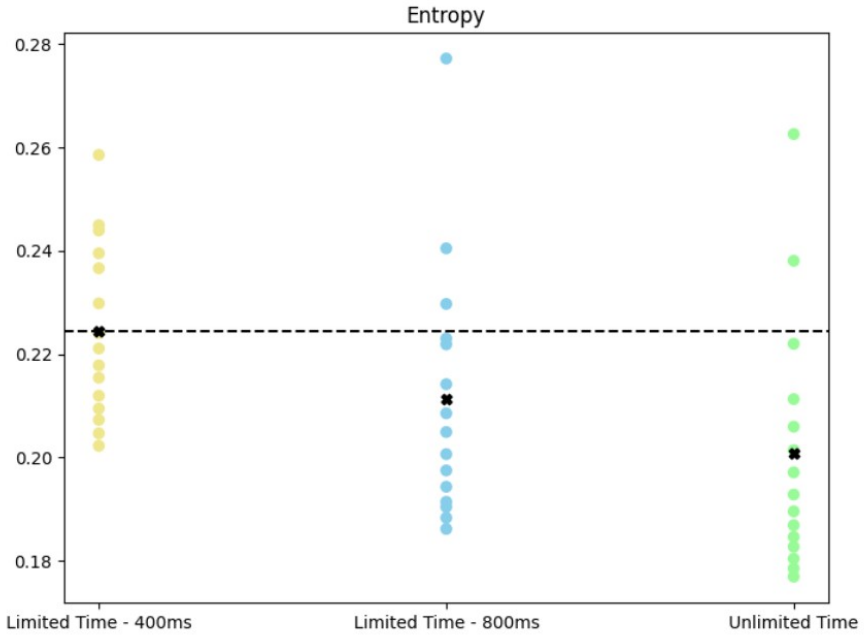


This plot shows the mapping of rewards to the keys q, w, o and p in the various payoff conditions as obtained from the participants' data. The mean and variance of the reward distributions are indicated in the plot.

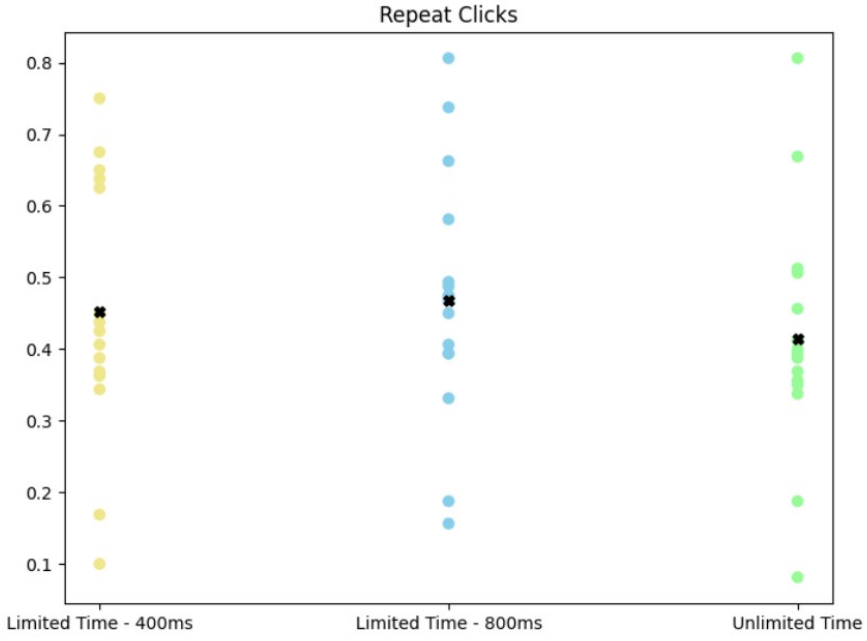


For all payoff conditions, the overall performance in the case of a limited time of 400ms seems to be lower while remaining almost for the other two time conditions. Incase of high variance and IGT conditions, there seems to be large fluctuations in the performance while in the low variance

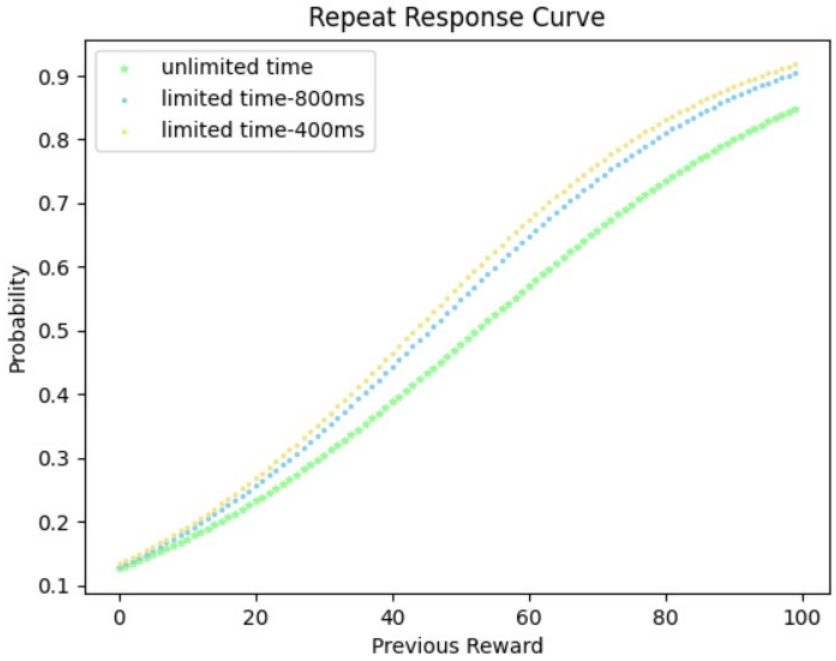
conditions and equal means contion, there is less fluctuation. This could indicate more exploratory decisions under uncertainty.



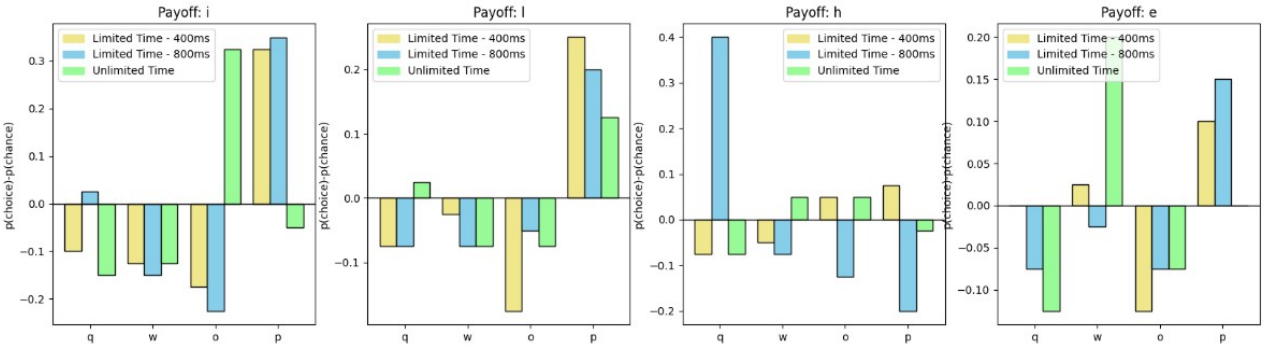
Here, the entropy of choices seems to be decreasing as the time pressure reduces. This implies that time pressure results in increased exploration in this case. This could be due to more risk-seeking decisions or the urge to choose more impulsive exploratory options under time pressure.



Here we observe that overall the amount of repeat clicks is lesser for unlimited timed rounds compared to timed rounds. This could imply that in the absence of time pressure, more exploratory behaviour is observed and thus the number of repeats are fewer.



When plotting the repeat choices as a function of the previous reward, we find that people were more likely to repeat their choices if the previous rewards were higher in the case of limited time conditions.



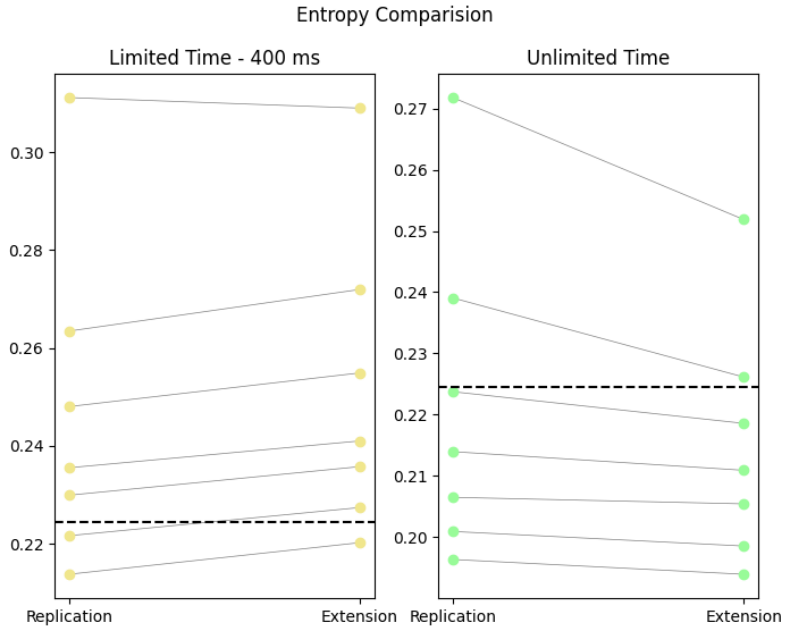
Lastly, we plot the aggregate choice frequency for each of the options under different payoff conditions. We find that option O (high reward low variance) was the most frequently chosen option in IGT condition in in unlimited time conditions but option P (high reward, high variance) was chosen most during both the timed rounds. This could imply a more exploratory behaviour in the lack of time pressure. In equal means condition, it is observed that option W, a low reward option, is chosen frequently in unlimited time conditions. However, in limited time, they were more inclined to option P which has high variance or uncertainty, indicating that when there is time pressure, people were more risk averse.

3.3 Comparative Analysis

The study of context is done by comparing the 400ms time pressure and unlimited time condition rounds of the same participant in the replication and extension experiments. No other parameter is changed in the two situations, and it represents the accurate effect of a contextual scenario while making a decision. A hospital scene is presented in the extension game with the stimuli of four tools. An updating graph to increase the stakes is introduced. The effect of this context is studied in three fields.

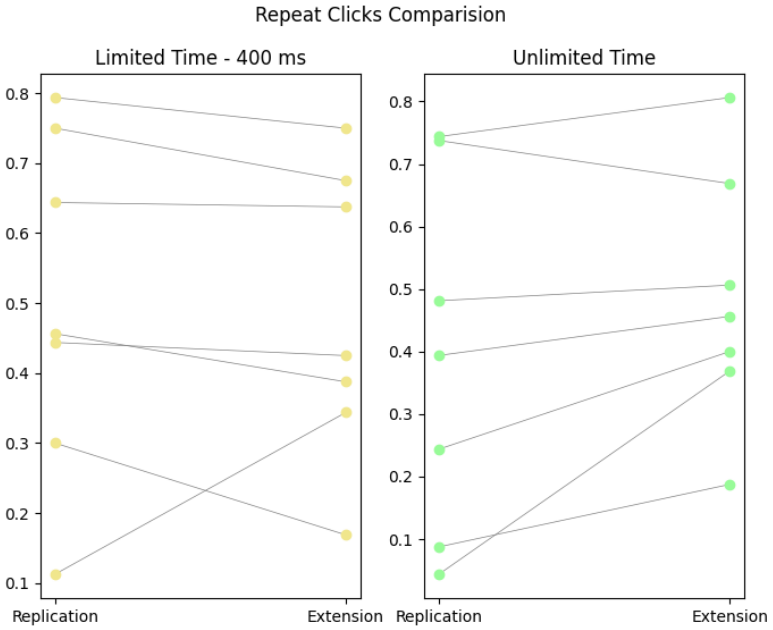
Out of the 15 data sets, we have collected 7 data sets for both games played by the same participants. Their performance in terms of entropy, repeat clicks, and response time is studied with respect to context. This shows interesting results with opposing effects on limited and unlimited time conditions, as detailed below. Each dot represents a participant's performance in the two games, which are joined with a grey line to show a direct effect seen on each participant.

3.3.1 Plot 1: Entropy vs. Context



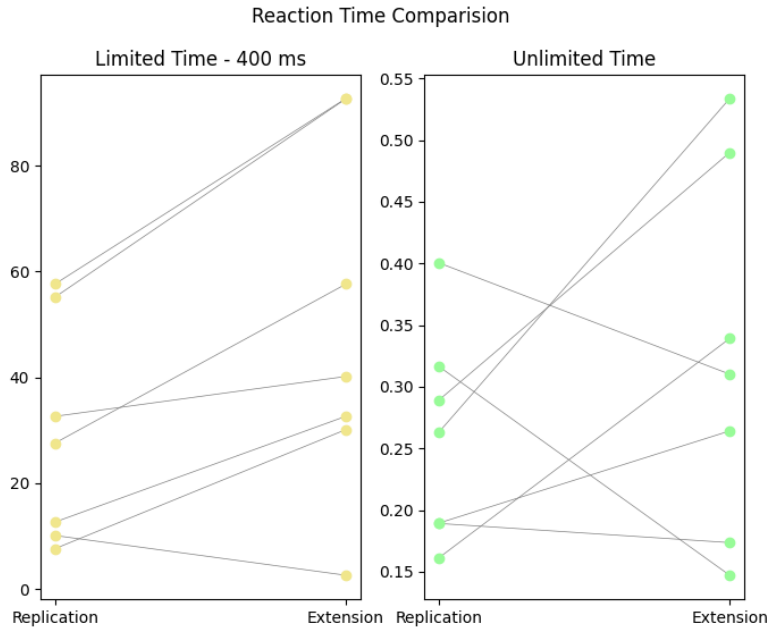
The entropy is seen to increase in six out of seven participants, with a noticeable margin in the limited-time case. The instinctual nature is seen to be further catalyzed when the context is introduced, leading to increased entropy values. On the other hand, in the unlimited time duration, the effect is the opposite as it decreases the entropy in all seven participants. More value-based decisions and lesser random choices are observed, leading to lower entropy values.

3.3.2 Plot 2: Repeat Clicks vs. Context



The Repeat clicks are plotted against context to observe its effects. In the limited time conditions, a decrease in the repetition of clicks is observed in six out of seven participants. This result is consistent with the instinctual approach explained by the previous graph leading to more random choices. In the unlimited time case, six out of seven reported an increase in repeated clicks showing a more exploitative approach taken when a looming medical context is introduced.

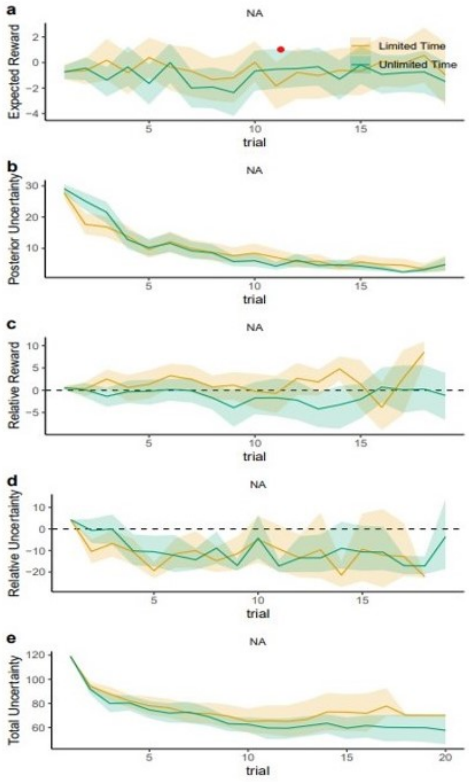
3.3.3 Plot 3: Reaction Time vs. Context



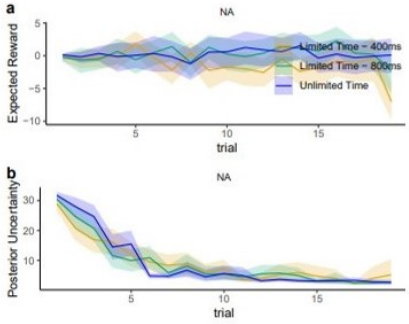
The Reaction time is observed in relation to the introduced hospital scenario. In the limited-time duration case, an increased reaction time is observed for the extension game for six out of seven participants. A longer computational time is demonstrated through this plot with the increasing stakes and stimuli when a hospital scenario is presented. In the unlimited time case, there is no observable pattern showing the insignificance of the reaction time of a decision process when there is no limit placed on it to induce pressure.

4 Conclusion and Discussion

We generated figures corresponding to 1c, 2a, 2b,2c,2d,2e from the paper for both replication and extension game. Additionally, we had also compared entropy, Repeat clicks and Reaction time for both the replication and extension task. We had also tried generating model-based predictions using the codes that came along with the paper.



Replication



Extension

BMT predictions about the chosen option simulated for all participants

We obtained a similar profile for replication fig b, d, e, and extension fig a and b corresponding to fig S5 in the paper. However, the replication fig a and c showed a varied pattern. Two likely reasons are:

1. Lack of data points available to us.
2. Mental state of the participants (nervousness, anxiety, seriousness, etc.)

Critique:

1. The paper showed data only corresponding to Q, W, O, P mapped with 1,2,3,4 options and not other combinations for each payoff condition.

Overall, the paper helps us in understanding how time pressure changes the way people behave and respond to uncertainty.