COMPARING HUMANS TO MICE: SIMILARITIES AND DIFFERENCES IN A MATCHING TASK WITH DELAY MANIPULATIONS

BACHELOR'S THESIS

BSC COGNITIVE SCIENCE

Maris Sala *Author*

Joshua C. Skewes *Supervisor*

Juliane Martin *Co-supervisor*

December 2018

Aarhus University

TABLE OF CONTENTS

SUMMARY

According to the generalized matching law, an agent matches their choices to the experienced reward rates of the choice alternatives (Herrnstein, 1961). A dynamic probabilistic foraging task revealed that if the mouse had to wait longer before making their choice between left and right side, they were observed to obey the matching law better, switched less between the two sides and integrated their past choices less than in a short delay. To compare, a similar task in the form of a simple computer game was conducted in humans with an added questionnaire about the task.

No difference of wait delay on matching in the human participants was found. In contrast to mice, past choices had a similar effect in all delay conditions. Questionnaire data revealed that people who reported their goal to have been to collect as many points as possible demonstrated a more exploitative pattern in their individual sessions. People who reported to have been trying to figure out the pattern showed effects of exploration in their trial-by-trial choices.

It was found that people's ability to identify and report the richer option throughout their experiment did not correlate with the side that gave them more rewards in the game. Thus, no hypothesized bias was found. Lastly, people's rating of their performance during the game was not reflected in matching better to the reward rates.

In conclusion, differences between humans and mice were found, mainly in delay effects being present or not. The human data was, however, noisier so it is difficult to say to which length the effect is observed. Future studies could test longer delays on humans and increase motivation with monetary rewards.

INTRODUCTION

Decision making has been studied in various experimental structures in animals and humans with an aim to understand common principles and in animals to understand the underlying neural mechanisms. To gain realistic insight, the experiments try to replicate real-life scenarios. The organization of the brains across mammalian species are largely comparable (Wallis, 2012), thus studying animals can give insight to human behavior and neural circuits if they're behavior correlates. Foraging behavior, which is a type of decision-making, can be studied in the framework of the matching law, according to which agents have been observed to match their choices between alternative options according to the experienced reward rates from the respective options (Rutledge et al, 2009). The experiment conducted in this paper additionally manipulates the delay before the decision between the alternatives can be executed. Choices and rewards are normally not evenly spaced in time which is what the delay manipulation will help investigate. The experiment is initially conducted in mice and the thesis is about testing the same experimental schedule in humans to see if there are behavioral differences.

1. THEORETICAL BACKGROUND

1.1. VALUE-BASED DECISION-MAKING

The two main types of decision-making are perceptual and value-based decision-making. Perceptual decision-making is about using sensory information to judge between stimuli, like which object is closer in space to the viewer. In contrast, value-based decision-making is subjective and depends on internal values like preferences, past experiences, and context – what alternatives are available, social effects and so on. A value-based decision would be deciding where to go for lunch (Rangel, Camerer and Montague, 2008). Studying value-based decision-making can give insight to the beforementioned internal states related to making the best choice in these situations.

An example of value-based decision-making are the choices made in foraging. Foraging is the act of finding and exploiting food resources, or more broadly any kind of resources necessary for life and reproduction. Foraging can be seen in the explore-exploit framework where in order to discover better resources one must explore, but to maximize rewards from the resources, one must exploit. The key is to find a good balance between exploring and exploiting. Some relevant factors for this balance are how well the agent knows the environment and the alternative options, how volatile these options are, and the cost of reducing uncertainty about the alternatives via exploration compared to exploiting known sources of reward (Cohen, McClure, & Yu, 2007).

In decision-making experiments where agents need to choose between options that give reward with a different probability, it is observed that given that the agents both explore and exploit, they are sensitive to these reward probabilities and tend to adapt their choice behavior to the task schedule (Herrnstein, 1961; Sugrue, Corrado, and Newsome, 2004; Rutledge et al, 2009). Reinforcement schedules are such task schedules where a chosen rule determines how rewards are distributed in a task. The relevant reinforcement schedule in this thesis is the variable-interval (VI) schedule in which a reinforcement is given at varying time intervals after the choices, but the average reward rate is kept constant. In an experiment, this can be achieved via assigning the choice alternatives specific reward ratios (probability). A VI VI schedule additionally assumes that these reinforcement ratios are independent for the two options in the task. If one side gives reward 10% of the times and another

40%, then a good explore-exploit balance would be to visit one side once and the other four times before switching to the first side again.

Another way to describe that type of behavior is via the generalized matching law, which states that agents tend to match their decisions between two choice options based on the received reward rates of these choice options across an experimental session (Herrnstein, 1961). The law is described by the following formula:

$$\frac{C_R}{C_L} = \frac{Rf_R}{Rf_L}$$

where C_R corresponds to response rate to right side, C_L is the response rate to the left, Rf_R and Rf_L describe the rates of reinforcements to the corresponding sides. In normal conditions, animals don't match to reinforcement rates perfectly, instead they have been seen to undermatch (Lau & Glimcher, 2005).

1.2. DECISION-MAKING IN ANIMALS AND HUMANS

Experiments on the matching law have been made in animals and humans. Results from animal tasks can be beneficial for various reasons. Animals can be trained to become experts in specific tasks and larger quantities of data can be collected. More invasive single-cell recordings can further be used to retrieve neural information that would correlate with observed behavior. Gaining insight to animals' internal states can correlate with human's internal states if their behavior matches (Wallis, 2012).

Sugrue, Corrado and Newsome (2004) used a similar dynamic foraging task in rhesus monkeys, where they needed to choose between two visual targets. They found monkeys to be reward-sensitive and able to adapt to the changes of reward probabilities within an experimental session. Lau and Glimcher (2005) also studied this in rhesus monkeys where they found the monkeys to undermatch to the reward rates. They're reason for observing a stronger undermatching is that they changed their reinforcement schedule many times in an experimental session, this has been observed also by other researchers (Davison & Baum, 2000; Palya & Allan, 2003).

Rutledge et al (2009) conducted a dynamic foraging task in humans. They studied the difference in behavior between people with Parkinson's on or off dopaminergic medication, matched with healthy elderly and young controls. They found that all groups' choice behavior changed according to the reward ratio changes in the experimental session on the group-level.

In conclusion, it's seen in both animals and humans that their choices depend on the reward ratios experienced in the task, meaning that they obey the matching law.

1.3. MATCHING IN MICE

A dynamic foraging task was conducted in mice by Juliane Martin in Kvitsiani lab (DANDRITE, Aarhus University). The experiment followed the same structure as the one done by Rutledge et al (2009). To introduce a more similar time variance present in nature, a wait delay was introduced in the initiation phase of each trial. Two delay intervals were used: a short (0-0.5s) and long (2-2.5s) delay. It was observed that the mice undermatched in the short delay condition but allocated their choices to reward ratios better in the long delay condition. A logistic regression analysis of past choice and reward effects on the current choice showed that the effect was seen due to differences in integration of past choice information, and not reward information. Thus, the animals didn't become more reward-sensitive in the long delay condition even though the matching initially suggested that.

1.4. AIM OF THE STUDY

The aim of this study is to conduct a similar experiment in humans with the wait delays to see if a similar effect is observed. To check if individuals' behavior can be grouped based on common differences, individual visualizations are also conducted. The participants were asked additional questions about their experience. The reported answers will be related to the behavioral data.

The research question for the behavioral data of the humans is as follows:

> Does the wait delay induce similar differences in matching behavior in humans as it does in mice?

The specific hypothesis related to the questionnaire are as follows:

1. People notice that the wait delay and reward ratios change throughout the experiment;
2. People's adaptability to changing reward ratios depends on their goal and is related to explore-exploit trade off;
3. If people report one option to have been better than the alternative, a bias towards that side is expected to show in matching law;
4. People who rate their performance more highly should show a better matching to the reward ratios in the experiment.

2. METHODS

2.1. PARTICIPANTS

The experiment was distributed as a computer game to be downloaded from email. Before they could proceed to the data collection phase, they were asked to sign a consent form. The data was sent directly to the assigned receiving email and was not stored on the participants' computers. The experiment and questionnaire data were collected from 17 individuals (10 females, 4 males, 1 other). The participants reported an age group to which they belonged to: 8 were in the group 18-30, 6 were 31-40, 2 were 41-50, and one was 51-60 years old. 15 were right-handed and 2 were left-handed. In the experiment, 12 used both hands and 5 only used their right hand. Both left-handed people used both hands.

2.3. DESCRIPTION OF THE EXPERIMENT

This experiment is a dynamic foraging task with a VI VI schedule. The probabilities for the two sides were 10% and 40% and they were independent of each other. There are three wait delay conditions for the gate opening: small (0-1s), medium (2-3s) and long (4-5s) delay. The trials are organized as blocks, where each block has 20 to 30 trials. With two reward probabilities and three wait delays, there were six blocks. What makes the probabilities dynamic is that if a reward is assigned to a side, it stays there until it is collected which means that the probability of reward on the unvisited side rises in time.

The task is based on an experiment conducted in mice. Figure 1 shows a visualization of the mouse experiment schedule. The mouse initiates a trial by poking their nose in the center port and waiting for the length of the assigned wait delay. This is followed by the free choice state where they either poke to the left or right port. These ports have their respective probabilities of giving reward in the form of water (P_L for left and P_R for right side).

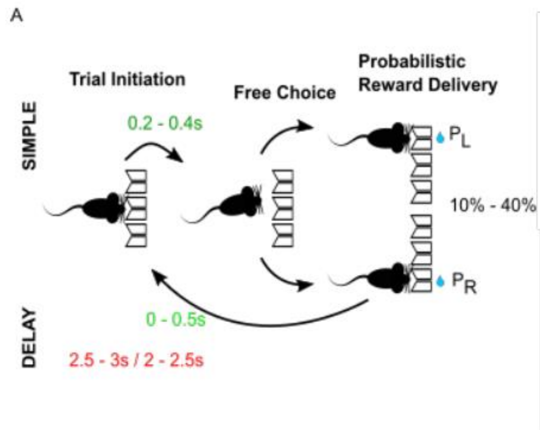


Figure 1. Overview of the mouse experiment. First, the mouse pokes his nose in the center port, there he waits for the length of the wait delay. Then he enters the free choice state and pokes either left or right port.

The human experiment was built as a computer game in Unity. The participants downloaded the game, went through some tutorial trials, and played the game. The participants are making decisions in the game as an avatar who is controlled with the "space bar" and "arrow" keys. Just like the in the animal task, the agent initiates a trial by waiting in front of the gate (Figure 2). To do this, the "space bar" needs to be held down until the gate opens. If the "space bar" is released before the wait delay is reached, the trial begins again until the agent waits sufficiently long.

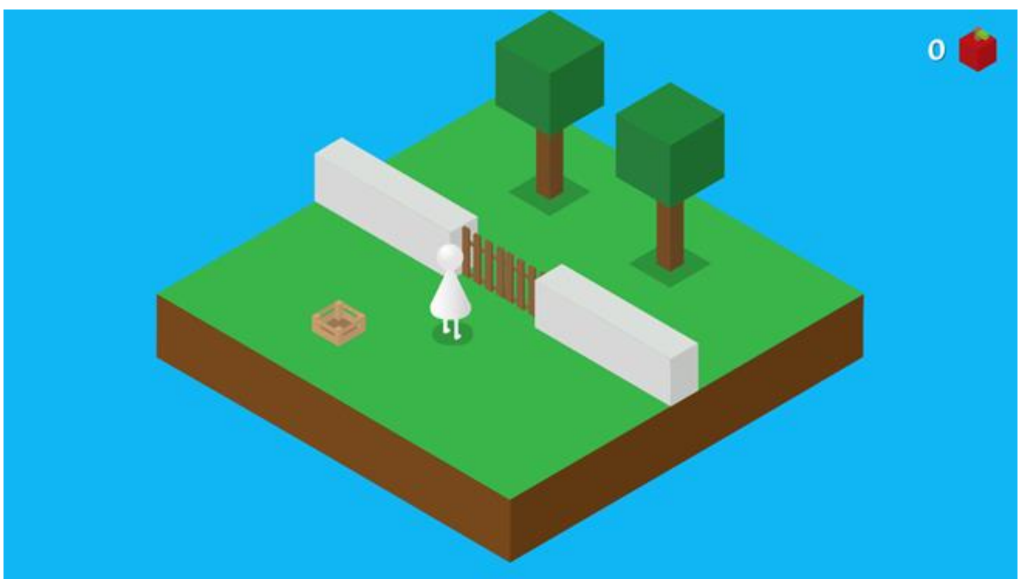


Figure 2. Screenshot from the game. The avatar waits in front of the gate until the "space bar" is released. Then it chooses between the right and left tree. If the chosen tree is rewarded, an apple will fall and will be taken to the wooden crate. On the top right corner there is an apple icon with a counter that shows how many apples the participant has collected.

Once the gate opens, the avatar walks to the trees and the participant chooses which tree to explore (free choice stage) by using the "arrow" keys. If there is a reward, an apple falls from the tree and the participant gets a point.

The participants play the game until they are notified that all necessary six blocks have been completed. After the game is completed, they are directed to a short questionnaire to assign a randomized ID to their data. It consists of questions about demographics (age group, gender, handedness). They were asked to report which hand(s) they used on the keyboard during the game, to select which of the given options was their goal and what they think changed during the game. They're asked to report which tree gave more reward in their experience, if any, and to rate how well they think they optimized for their goal on a scale of 1 to 5 where 1 is "very badly" and 5 "very well". The questions with the respective options of answers are reported in Appendix B.

2.4. DATA PREPROCESSING

The data was analyzed in RStudio version 1.0.153 with R version 3.4.1. All plots were made using ggplot2.

The experiment data was joined together with the questionnaire data. All participants data was included in the analysis. There was a mean of 162.4 (SD = 10.7) trials per person. Their attempts per trial at the gate were recorded. A first attempt is when they waited sufficiently on first try, further attempts were recorded until they waited for long enough.

The human data was analyzed similarly to the mouse data. Four methods were used to describe the behavioral data for both experiments. These include individual plots of choices during a session, run lengths, the matching law, and logistic regression of the effects of past choices and rewards on current choice.

To visualize individual's behavior, moving average of choices in a 10-trial window was calculated per participant and plotted. Choices towards right side are encoded as 1s and towards the left are encoded as 0s. The moving average of choosing right side in the trial window is calculated and is to be interpreted as the probability of choosing right side over the left side.

Run lengths show how many times in a row did the agent choose a side before switching to the alternative and they can give additional insight into behavior. Right choices are encoded as 1s and left choices as 0s. Run lengths are visualized as frequency plots.

The matching law describes how well the agents match their choice rates to the reward rates experienced in the experiment. For the matching law, right choices and rewards were encoded as 1s, left choices and rewards as -1s and non-rewarded trials as 0s. For each participant, a sum of choices and rewards to the respective sides were calculated per block thus there are a maximum of 6 data points per person since there are 6 blocks. Then the logarithm of the ratios between right and left choices and rewards were plotted.

However, the logarithm of 0 is not defined, thus if no rewards were collected in an entire block, that data point was discarded, although collecting no rewards would conceptually still be meaningful. This is a common practice when calculating data points for the matching law (Herrnstein, 1961; Sugrue, Corrado, and Newsome, 2004; Rutledge et al, 2009). Accordingly, 0.1% of the data was excluded in the matching law calculations. There was one participant for whom 2 out of 6 data points were lost.

To describe the effects of past choices and rewards on current choice, a general linear model was composed using binomial family and the logit link function. Right choices were encoded as 1s and left choices as 0s. Right rewards were 1s, left rewards -1s and non-rewarded trials were 0s.

Current choice was predicted from past 20 choices and rewards. To shrink the coefficients and control for correlations between choices and rewards, ridge regression was used.

$$y_{choice} \sim \beta_i + \beta_1 C_{t-i} + \beta_R R_{t-i} + \beta_L R_{t-i} + \varepsilon$$

The coefficients were plotted against t - i trials and a line was fitted using loess fit.

3. RESULTS

The behavior of the human participants was visualized and compared to the animal experiment results.

Visualizations are made per person and across all participants to see the average effects. Individual plots give insight into how people behaved individually, and this could perhaps also indicate some common patterns in case there are groups of individuals who behave similarly.

All plots that distinguish between delay conditions use the same color coding: blue stands for the short delay, purple for medium, and red for long delay.

3.1. INDIVIDUAL SESSIONS

We see in animals that they adapt to the changes in the environment. If the reward probabilities change, the animals' behavior changes to choose the new more rewarding site over the older one. Rutledge (2009) found that humans, too, change choices based on reward changes. To check this, individual sessions were plotted per participant.

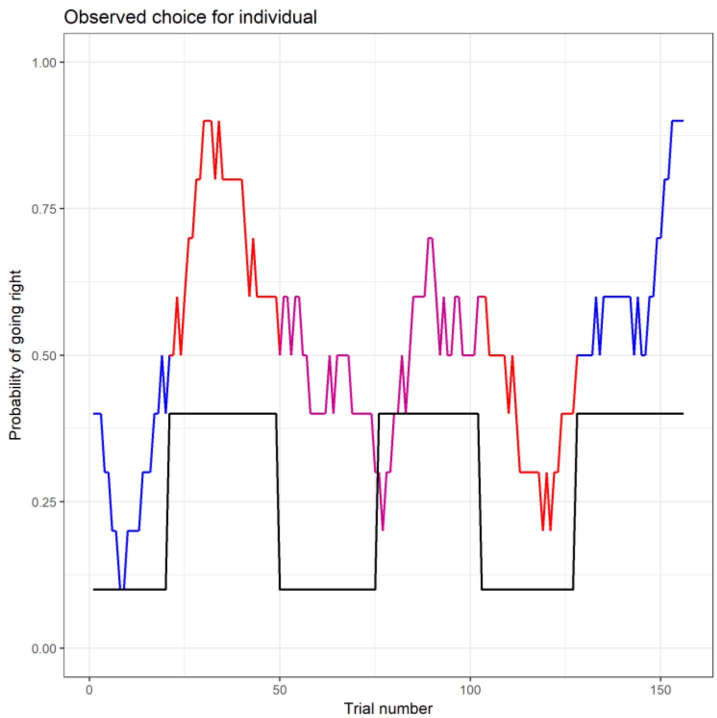
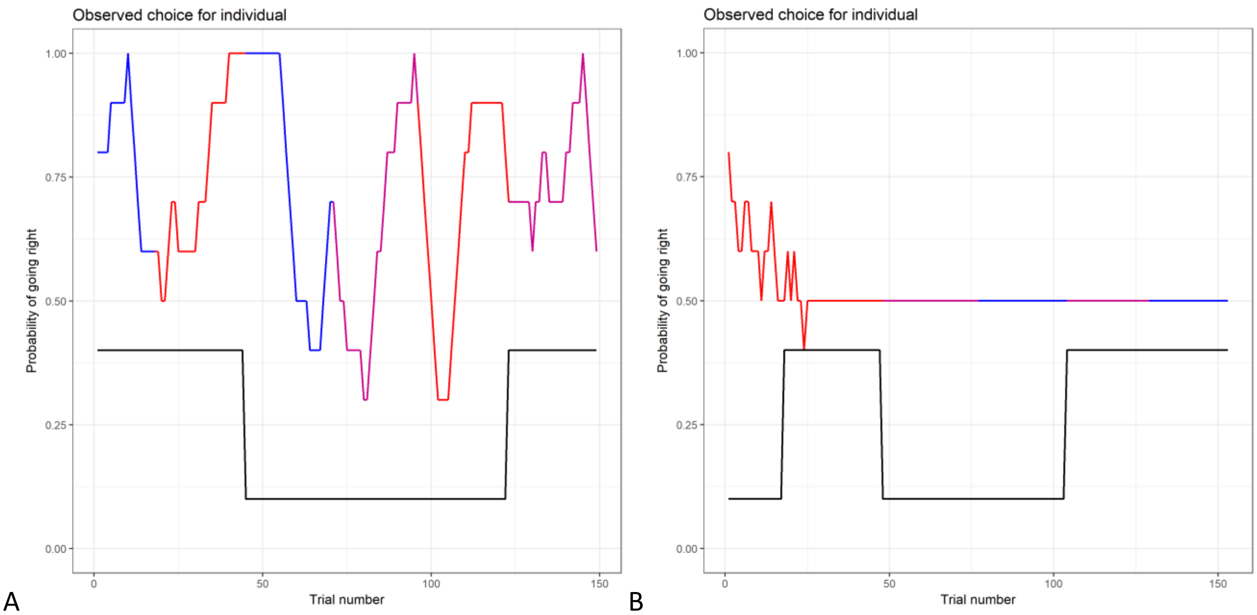


Figure 3. Observed choice of individual. X-axis shows the trial number starting from the beginning of their session until the end, y-axis shows the average probability of choosing right side over left side. The delay conditions are colored from blue to red, blue being the shortest delay and red the longest. The black line shows the probability of reward on right side.

The individual sessions are plotted as a moving average of choices during the session per person. Observed choice shows the relationship between trial number (x-axis) and the average probability of choosing right side over left side (y-axis). The black line shows the probability of reward on right side (10% or 40%). Overall, this individual's choices are affected by the probability of rewards – when the probability of reward is 40% then the agent has a higher chance of choosing right side.

Some individuals adapt to the reward probabilities better than others. In Figure 4, individual A's choices seem to fluctuate in the different reward probability conditions; furthermore, they seem to have a high average probability of choosing right side over left – a potential right bias is observed. For individual B, it's visible that they began with exploring the two alternatives, but after around 25 trials began to alternate between left and right, never choosing one side more over the other.



A B

Figure 4. Observed choice for two different individuals. A - a right bias is observed. B - a pure alternation between two sides is observed after the first 25 trials. The delay conditions are colored from blue to red, blue being the shortest delay and red the longest.

3.2. RUN LENGTHS

To better understand switching behavior, run lengths were calculated and visualized. Run lengths show how many times in a row did the agent choose one side before they switched to the other. It was observed that mice would switch more often in the short delay condition and stay longer on one side in the long delay condition.

Individual humans showed to have differing run length patterns. Some were like the mice, others were opposites (see Appendix C). Individual differences were too great to group them for comparison. To check for what humans did on average, their run lengths were added together and visualized. In contrast to the mice, the average behavior of the human participants doesn't show a difference in the three delay conditions (Figure 5).

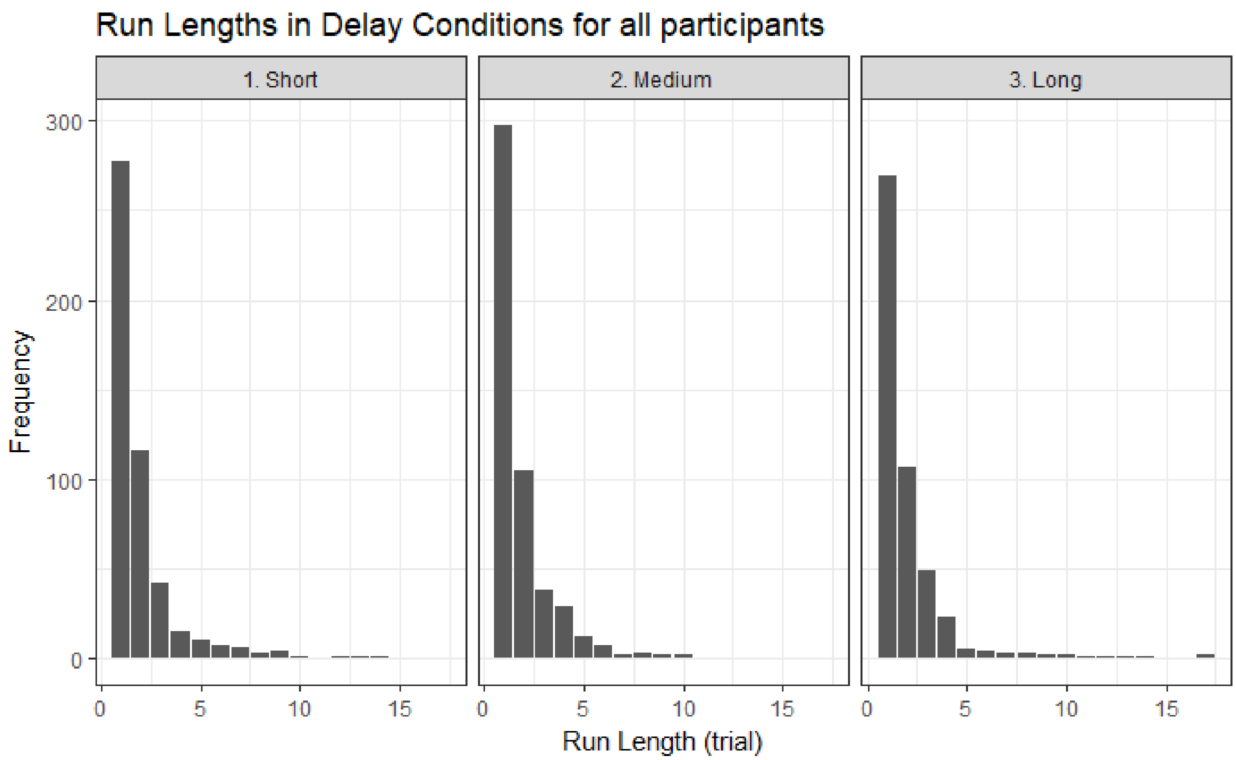


Figure 5. Run lengths for all participants in the three delay conditions. X-axis shows how many trials the agent chose to be on the same side in a row, y-axis describes the frequency of these choices.

3.3. THE MATCHING LAW

Visualizing matching in mice has shown that mice seem to match better to reinforcement rates when they are in the longer delay condition. In the short delay they tend to switch more often between the sides.

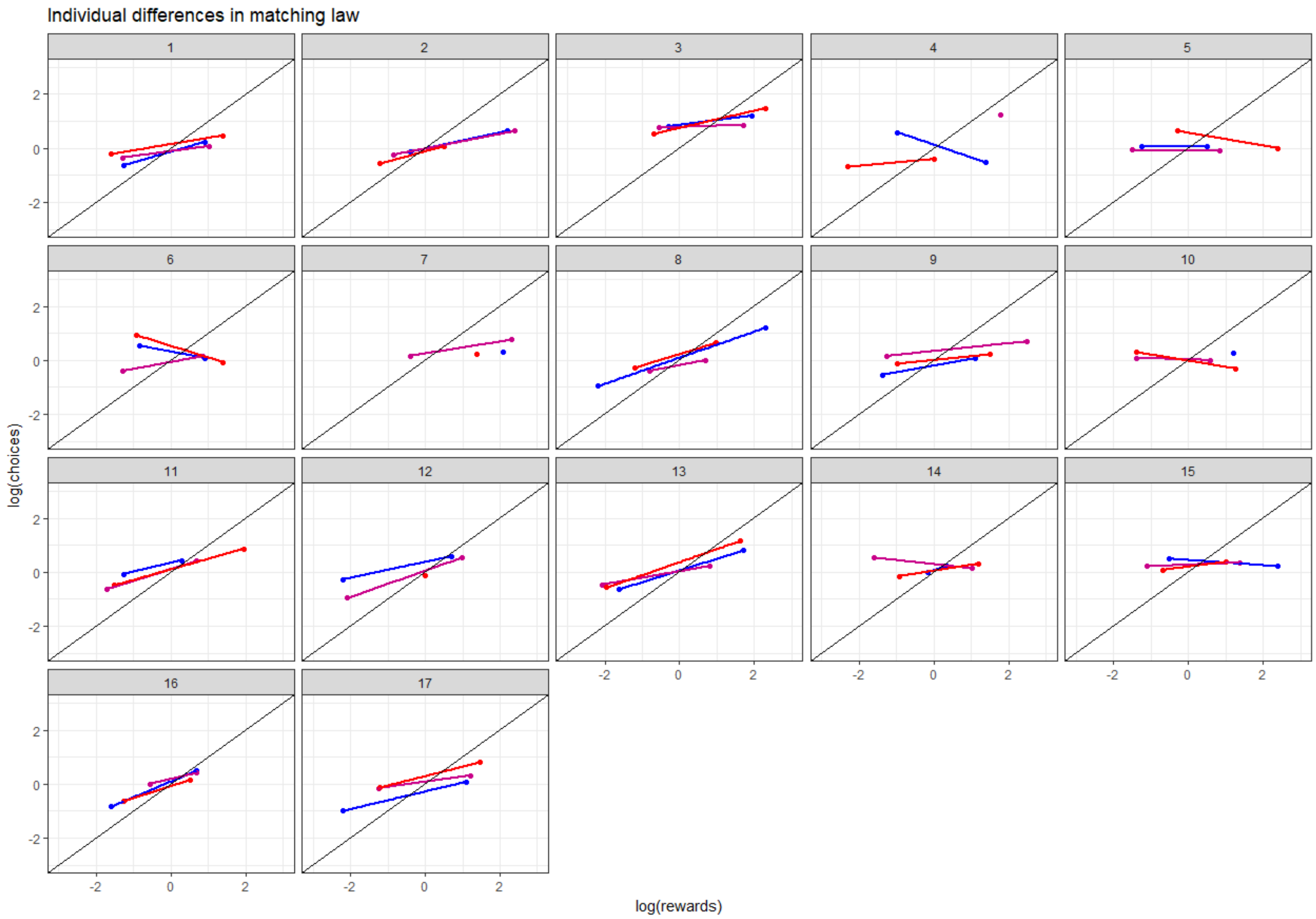


Figure 6. Individual differences in matching law per person. Each delay condition has two data points and a line has been fitted where possible. The delay conditions are colored from blue to red, blue being the shortest delay and red the longest. The diagonal black line shows the behavior of a perfectly matching agent. The closer the slopes in the data resemble the black line, the better they match.

The logarithms of choice and reward ratios were plotted, and the delay conditions were separated by fitting a different line for each, as seen in Figure 6 and Figure 7. The black line represents an agent who matches their choices perfectly to the reward ratios and the better the human data fits this line, the better they are at matching. Individually, most people seem to match but the data is noisy, and some people in certain conditions even do the opposite of matching (Figure 6, plot 4, blue - slowest delay). Since there are only up to 30 trials per block and the reward probabilities are below average for both sides, some people might have had “bad luck” in exploring the two sides, resulting in below-average matching.

On average, people do seem to match their choices to the reinforcement rates, but in comparison to mice, no visible difference between delay conditions is seen (Figure 7).

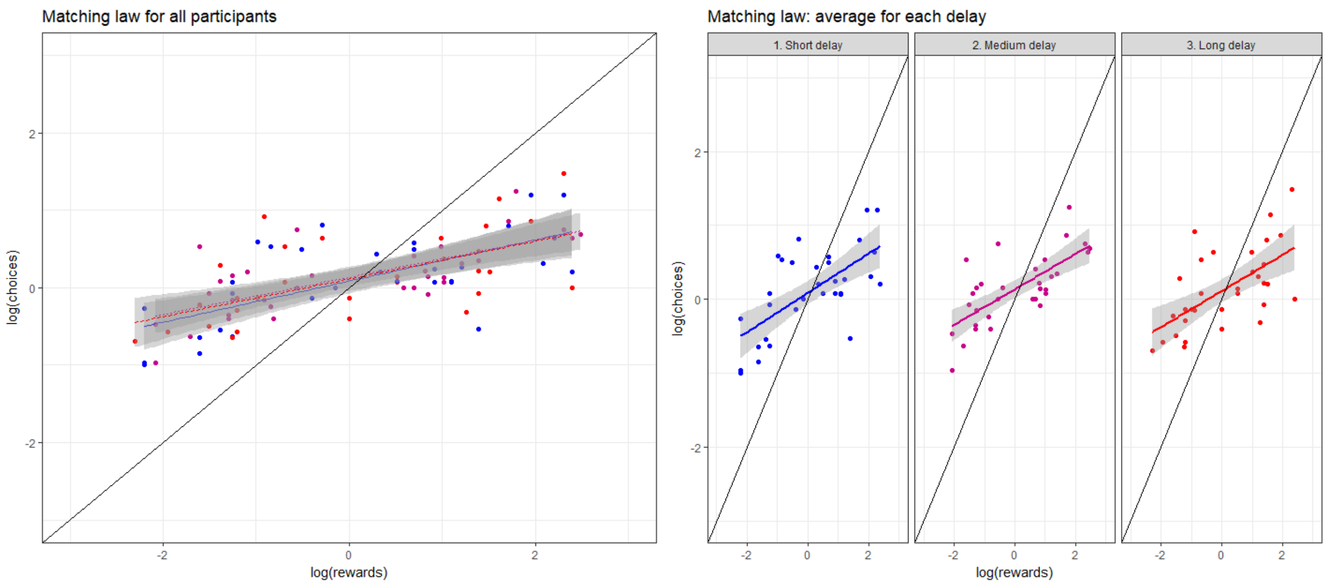


Figure 7. Log plots of matching law across all participants in the different delay conditions. The delay conditions are colored from blue to red, blue being the shortest delay and red the longest. X- axis shows the log of rewards on left and right sides, y-axis shows the log of choices to the same sides. The negative side corresponds to left side, positive side corresponds to the right side. Diagonal line shows the slope the agents would follow, if they would match perfectly. Left – All people and one line for delay condition. Right – Delay conditions plotted separately, for clarity.

3.4. LOGISTIC REGRESSION

To investigate why mice would match better in the long delay condition, a logistic regression of the effects of past choices and rewards to current choice was made. If the agents become more reward-sensitive in the long delay, changes in reward history between the delays is expected. If instead this effect is due to choice effects, then a change in choice history patterns in the delay conditions is expected.

For the mice, the past reward effects remain the same. However, past choices show a difference for short and long delay. The first coefficient in the short delay condition is more negative than it is in the long delay. This suggests more switching in the short delay, which has already been observed in the previous parts of analysis. This means that the mice are just as reward-sensitive in shorter delays, but their integration of past choices into the decision process decreases on long delays.

The past choice and reward effects for the human participants are visualized in Figure 8. The first negative coefficient describes a switch to the other side. Based on past choice effects, the agent tends to switch from current side to the alternative. Thereafter the coefficients become slightly positive and decay in time.

There is a positive effect of staying on the current side in both right and left rewards effects. This means that when a side has been rewarding, the agent tends to choose this side again next turn. Here, too, we see a decay towards 0 around 5 trials back from current trial. The confidence intervals are smaller for the first coefficients and larger later which could describe the differences in decay rates of past rewards in the different individuals.

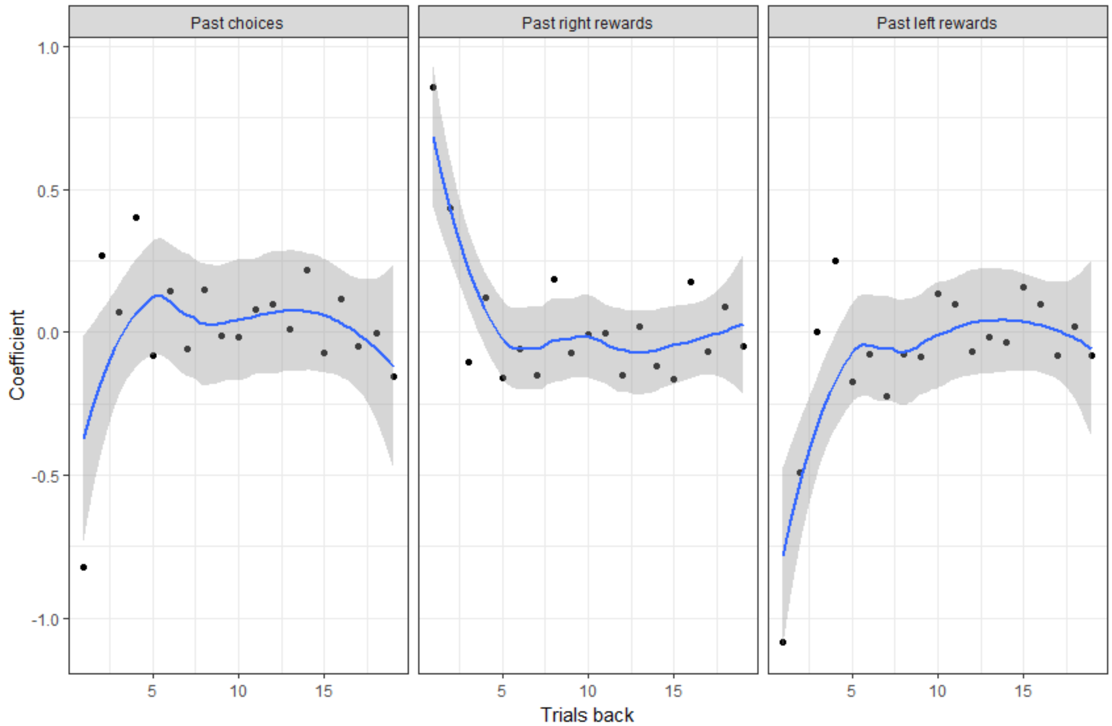


Figure 8. General linear ridge regression of the effects of past choices, right and left rewards on current trial.

To check for potentially similar differences in the delay conditions as was seen in mice, separate lines with confidence intervals for the three delays were produced. Figure 9 demonstrates that past rewards tend to have the same effect across delay conditions and the first coefficient seems to shrink in the longest delay.

The first coefficients for past choice effects don't seem to differ more than they do for past rewards. In delay 2 we observe the largest first negative condition, this might be due to noise. For both past choices and rewards, the longest delay has the lowest coefficient, then follows the short delay, and the biggest change is seen in the medium delay.

In conclusion, humans integrate past rewards and choices like animals with a difference of trend in past choice integration. Unlike in mice, an effect of delay condition couldn't be observed in humans due to higher levels of noise in the human data.

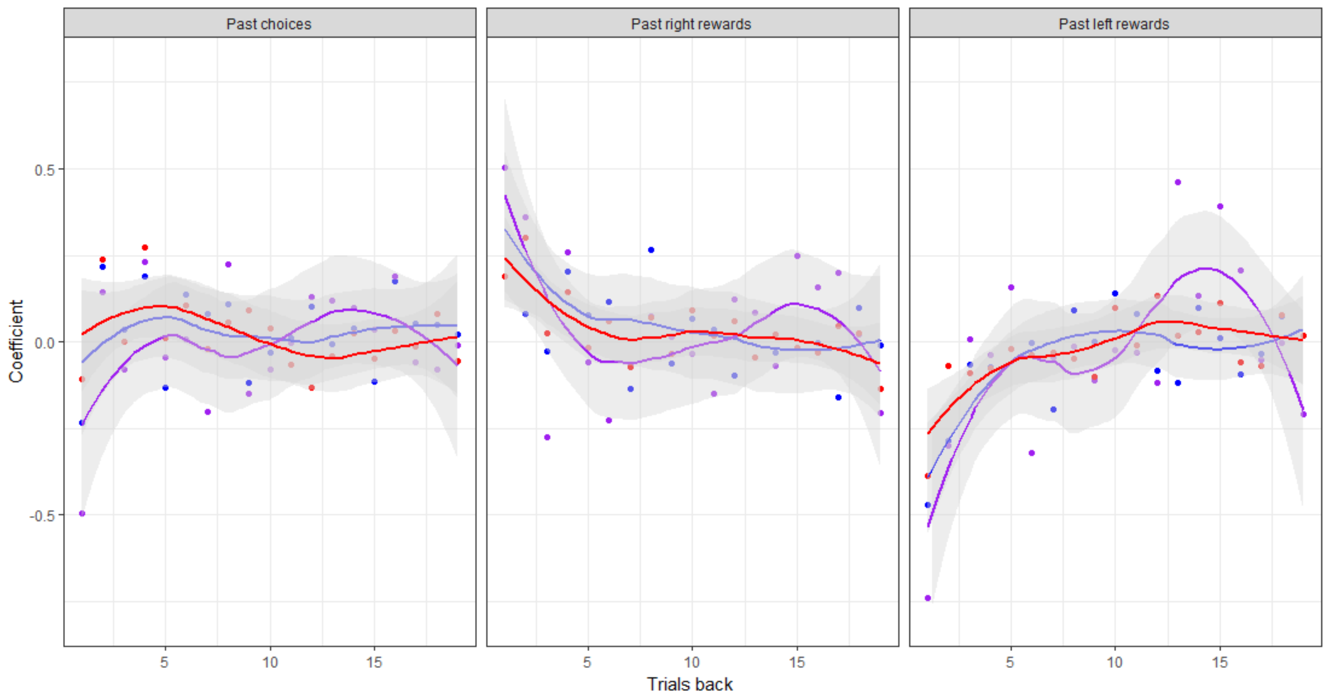


Figure 9. General linear ridge regression of the effects of past choices, right and left rewards on current trial in the different delay conditions. The delay conditions are colored from blue to red, blue being the shortest delay and red the longest.

3.5. RELATING QUESTIONNAIRE TO EXPERIMENTAL DATA

All participants answered some questions in the end of the experiment. Since the delay didn't prove to show a difference in previous analysis, the data across delays is concatenated in the following visualizations. The experiment-related questions were collected so that they could be related back to the behavioral data in terms of the matching law.

The participants were asked if they experienced a change in the following three conditions: wait delay and reward rate, and the walking speed of the avatar. The last was not manipulated in the task. All participants reported that they noticed the wait delay's changing and 14 people noticed reward rates changing. 6 reported that they noticed a change in the avatar's speed of walking.

3.5.1. Relating goal to the matching law

Out of 17 people, 12 reported their goal to be to collect as many apples as possible, 3 reported that they tried to detect the pattern of the rewards and two didn't report a goal. Comparing agents who aim to collect as many rewards as possible to the ones trying to detect a pattern of reward ratios is representable as a difference in explore-exploit balance. Exploitation is useful for collecting most rewards whereas exploration is needed for detecting the pattern (Cohen, 2007).

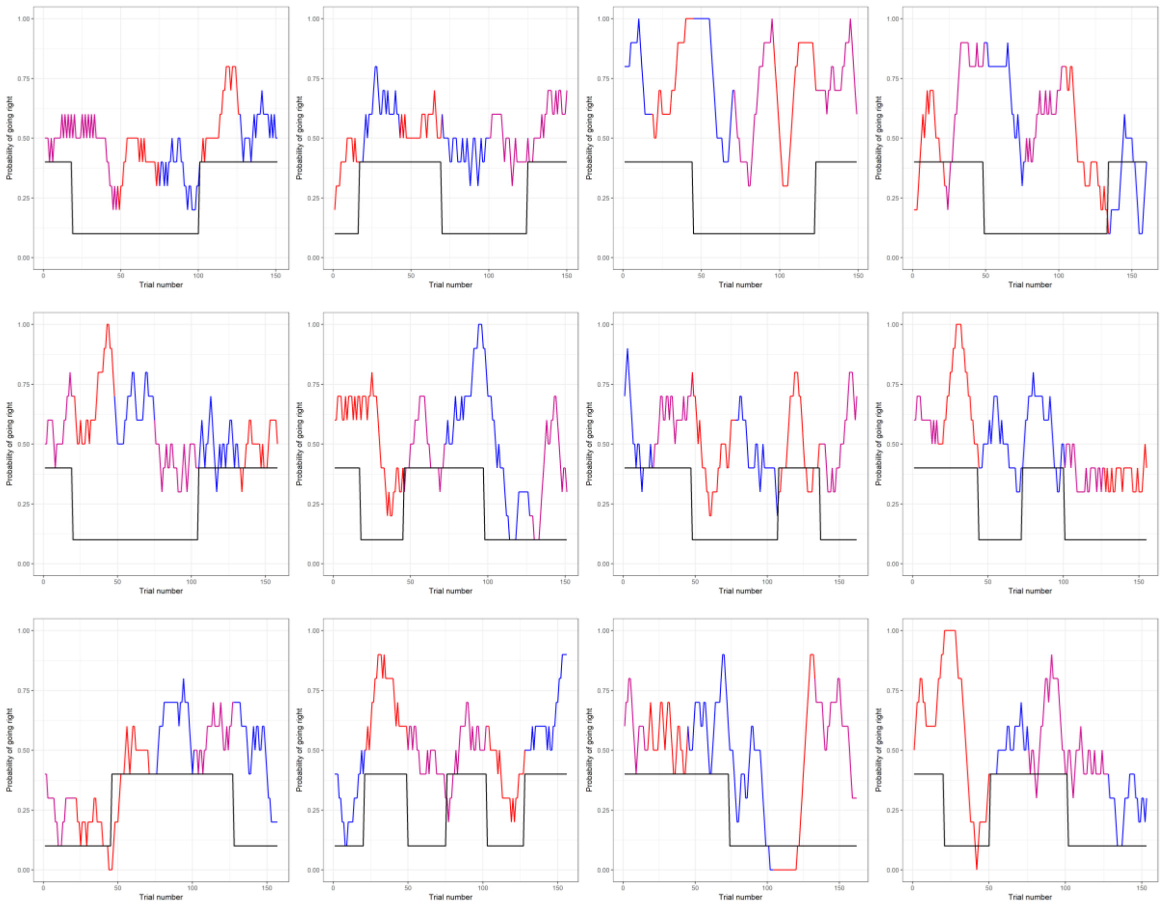


Figure 10. Observed choice for individuals whose goal was to collect most apples. The delay conditions are colored from blue to red, blue being the shortest delay and red the longest.

Of the methods used in this thesis, observed choice for individuals is best to investigate this question in a more qualitative manner. If they wish to collect most apples, they would be more sensitive to the reward ratio changes in order to maximize their exploitation from the richer option. If they want to detect the pattern, they would be expected to explore more from both sides, so their average choice probability to the right side should be around 50%. Figure 10 shows the individuals who aimed to collect most apples and they appear to adjust differently to the reward probabilities than the people whose behavior is visualized in Figure 11 where they were trying to find the pattern of rewards.

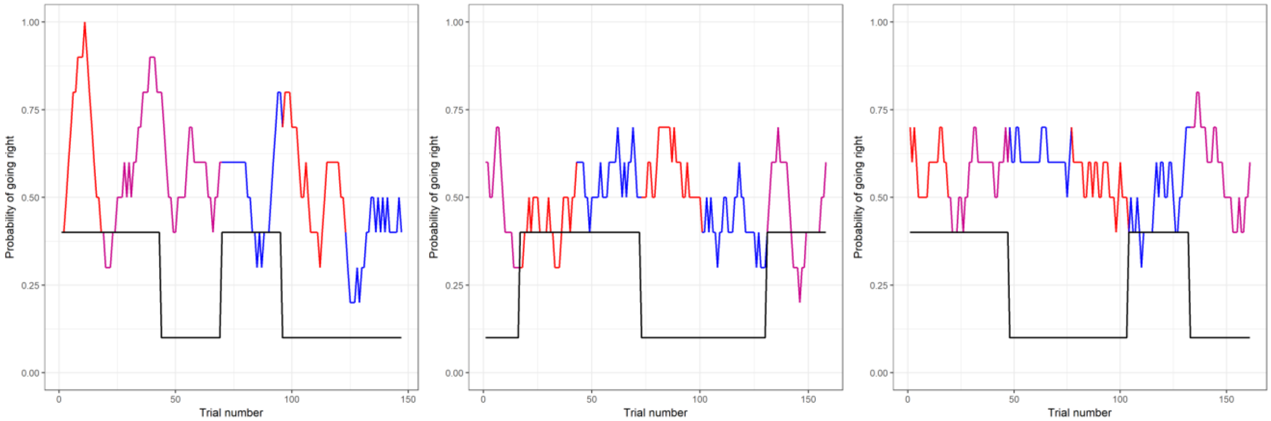


__Figure 11. Observed choice for individuals whose main goal was to detect the pattern of reward ratios. The delay conditions are colored from blue to red, blue being the shortest delay and red the longest.__

3.5.2. RELATING SUBJECTIVE EXPERIENCE OF THE RICH OPTION TO THE BEHAVIORAL DATA

Of 17 people, 7 reported that they experienced more rewards from right side, 2 from left side, and 6 report that they didn't notice a difference. Two didn't answer this question. Figure 12A demonstrates that most of the participants didn't report the actual reward difference (right-left). If there would be a bias to a side, it would be observable from the matching law plots (more choices would have been made to the side they believed was best). Since there wasn't a correlation between reported sides and actual sides, the matching law proved to show no meaningful difference (Figure 12B).

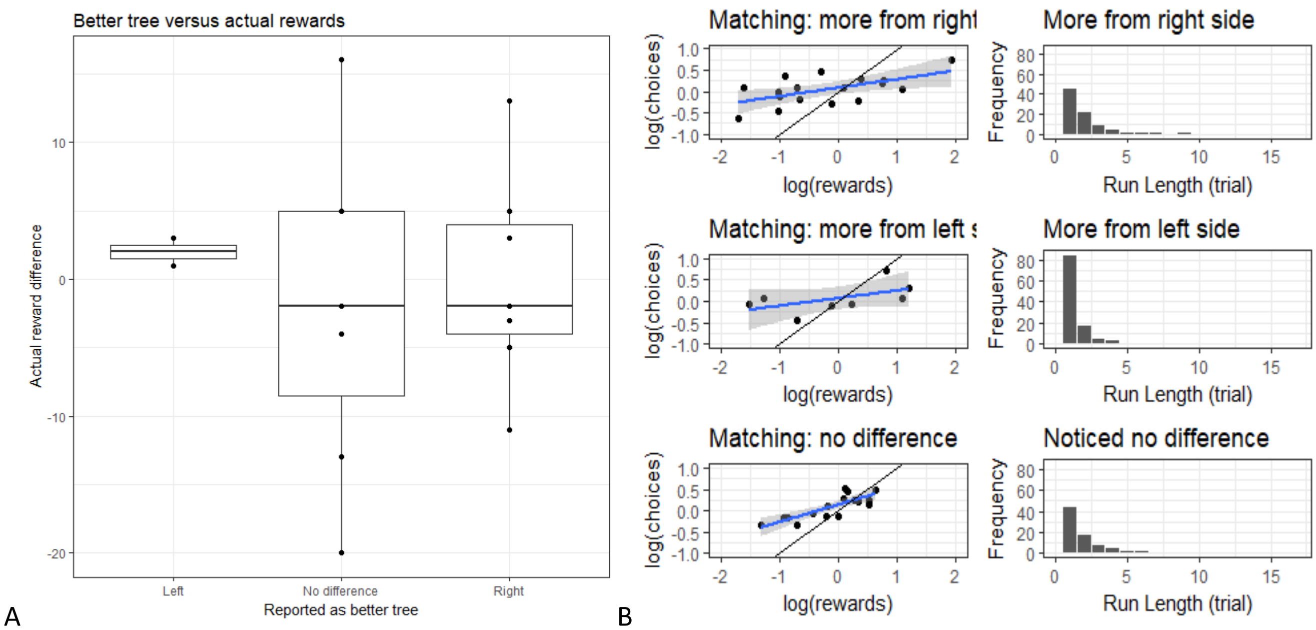


Figure 12. A - Comparing reported experience to actual experience of received rewards. X-axis shows what the people reported was the best tree, y-axis shows the difference between their actual collected rewards in the experiment. If the difference is positive, they gathered more from right side; if it's negative, they gathered more from the left side. B – Comparing matching and run lengths based on the trees people reported to have been better. No bias difference is observed.

3.5.3. Relating rating of performance to the matching law

A higher rating of performance should yield a stronger matching to the reward ratios. A stronger matching means there is a steeper slope. From the 17 people, 4 reported a personal rating of "2", 8 a rating of "3", and 5 a rating of "4". Figure 13 shows that there is no noticeable difference between the slopes for the three ratings that were reported.

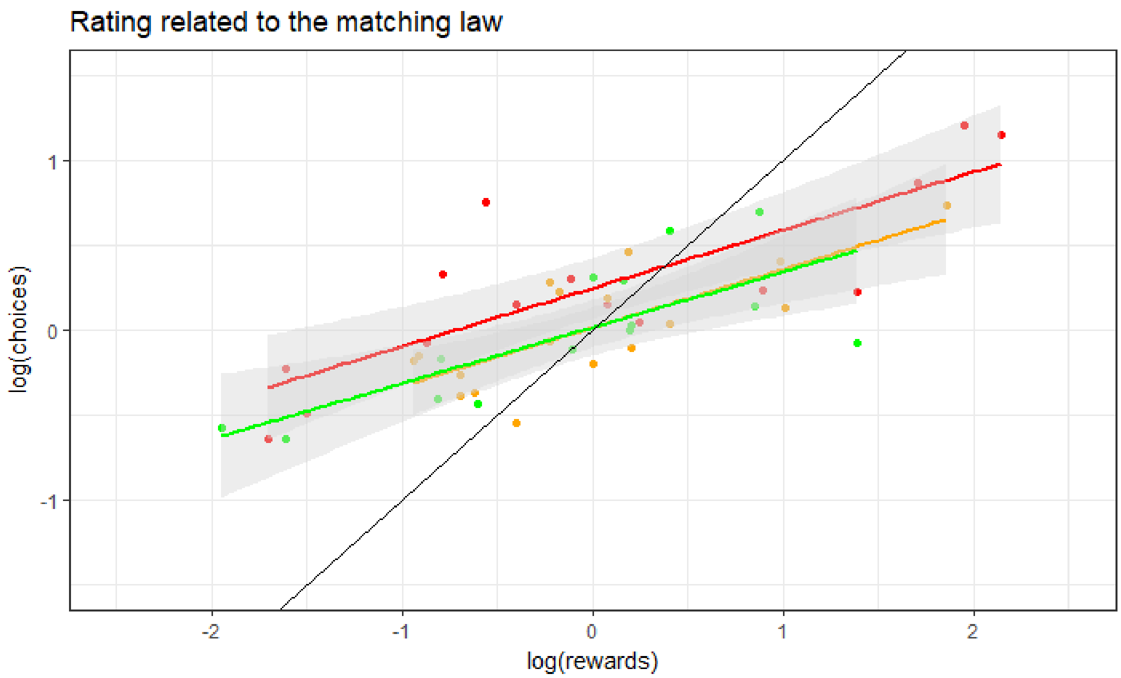


Figure 13. Relating matching law to the ratings people reported of their performance. Ratings went from 1 to 5 where 1 is "very bad" and 5 is "very good". 1 and 5 were not reported by anyone. Color coding of the ratings: red – 2, orange – 3, green – 4.

4. DISCUSSION

As seen in previous literature and the mouse task, we found that the humans in this experiment do adapt flexibly to the changes in the reward ratios during an experimental session. We did not see a difference in matching between the three delay conditions, in contrast to the mouse experiment. The logistic regression revealed that the effect of past rewards is consistent across the delay conditions for both mice and humans. The effect of past choices in humans doesn't change as much as it does in mice. Additionally, the first and second hypothesis about the questionnaire data proved to be correct – people noticed what changed during the experiment, but they also reported a change in the avatar's speed and this was not manipulated. The people might have experienced these changes, but they might have also reported these because they were among the answer options. A free answer field could have yielded different results. Their adaptability to reward ratios was observed to be influenced by their goal, although the two goals in this part of the analysis didn't have equal amounts of data. People were not good at reporting which tree gave them more rewards and they didn't match better to reward ratios when they reported a higher rating of their performance.

The delay could be linked to working memory span, so that more recent choices and rewards would matter more than the ones further back in time. Humans have a phonological loop span of 8-9s which means they can retain verbal information in the form of an implicit or explicit thought for almost 10s (Baddeley and Hitch, 1974). Since it is debatable whether mice have thoughts, their memory span might have a shorter span (Griffin and Speck, 2004). If the maximum delay in human experiment isn't longer than their phonological loop, then there might not be a difference between these conditions. It would be interesting to test if the mouse effects would arise in human participants if the delay would be prolonged to the phonological loop length.

Compared to humans, the mice are more motivated to participate in the experiment since they are water-deprived, and they receive water as a reward in the experiment. Human participants, however, receive points and no real rewards, they are also more likely to be better fed and rested when playing the game, thus their motivation is expected to be lower compared to the mice. In that case, the short delay might induce more anxiety because the agents need to decide faster, whereas in the long delay they have more time to process. It would be interesting to test this delay task in humans who have been diagnosed with anxiety or other diagnosis that affect cognition and mood. Instead of points, rewards could be of more relevant nature, like money.

Humans can maintain a strategy in mind while they are doing the task (Heracleous, 1998). Instead of trial-by-trial decision-making, people could already know which option they will choose next before the wait delay is over. In this case, manipulation of the delay would not make a difference to their choices. If mice operate less based on longer-term plans, then they might be more likely to decide in the end of the wait delay. That would instead indicate that they have a slower integration of past choices rate. However, past reward effects would then also be affected. It might make more sense for mice to integrate past choices better only if they have enough time for it and the short delay is too instantaneous to make choices matter. In contrast, people might not be affected by such slight context changes enough to drastically change strategies. Also, people might start doing that after having been trained for longer on this task.

Animals are used in decision-making experiments because they are easier to have access to and simpler to manipulate with. Animal experiments can give insight into human behavior and brain functions if their brains are comparable enough.

CONCLUSION

Where increasing delay before taking a decision yielded a difference in integrating past choices in mice, it did not yield a similar effect in humans. Their past choices were weighted similarly. People exploited more when their goal was to collect most points, and they showed a more explorative behavior when they reported to having tried to find the pattern of rewards. People were not good at relating experienced richer option across their session to the actual richer option and thus no difference in relation to side bias was not found in their matching. Lastly, people who reported a higher rating didn't match better in comparison to people who rated their performance less highly.

Future studies could design a task with longer wait delays for humans with more motivating rewards to see if the choice and reward effects are influenced by that. Additionally, both mouse and human data should be analyzed via fitting a model and since we see a behavior difference then it would be interesting to see if a different learning model would describe and predict humans better than mice. To find differences in questionnaire data, future experiments could give participants a goal in the beginning to see if and how their choices are impacted by that.

In conclusion, this thesis explored and compared the behavior of humans in a decision-making task previously studied in humans. Important differences in behavior were found and are interesting for future research.

ACKNOWLEDGEMENTS

I'd like to thank Juliane Martin and Duda Kvitsiani for the opportunity of working together with the animal and human data. I'd also like to thank Joshua Skewes for his time and insight.

BIBLIOGRAPHY

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Davison, M., & Baum, W. M. (2000). Choice in a variable environment: every reinforcer counts. *Journal of the Experimental Analysis of Behavior*, *74*(1), 1–24. https://doi.org/10.1901/jeab.2000.74-1

Griffin, D. R., & Speck, G. B. (2004). New evidence of animal consciousness. *Animal cognition*, *7*(1), 5-18.

Heracleous, L. (1998). Strategic thinking or strategic planning? *Long range planning*, *31*(3), 481-487.

Herrnstein, R. J. (1961). Relative and Absolute Strength of Response as a Function of Frequency of Reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*(3), 267–272.

Lau, B., & Glimcher, P. W. (2005). Dynamic Response-by-Response Models of Matching Behavior in Rhesus Monkeys. *Journal of the Experimental Analysis of Behavior*, *84*(3), 555–579. https://doi.org/10.1901/jeab.2005.110-04

Palya, W. L., & Allan, R. W. (2003). Dynamical concurrent schedules. *Journal of the Experimental Analysis of Behavior*, *79*(1), 1–20. https://doi.org/10.1901/jeab.2003.79-1

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature reviews neuroscience*, *9*(7), 545.

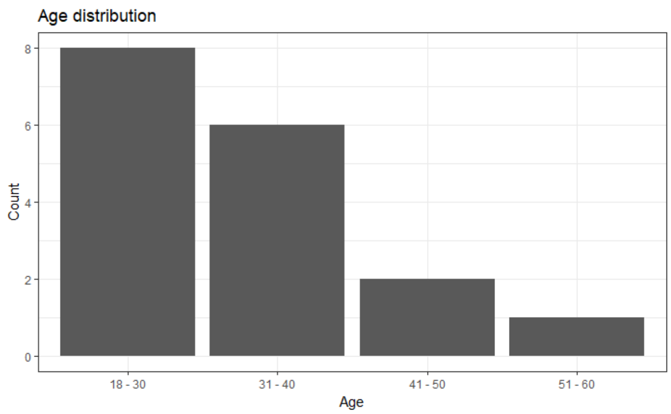
Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic Drugs Modulate Learning Rates and Perseveration in Parkinson's Patients in a Dynamic Foraging Task. *Journal of Neuroscience*, *29*(48), 15104–15114. https://doi.org/10.1523/JNEUROSCI.3524-09.2009

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching Behavior and the Representation of Value in the Parietal Cortex. *Science*, *304*(April), 66–74.

Wallis, J. D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature neuroscience*, *15*(1), 13.

APPENDIX

A



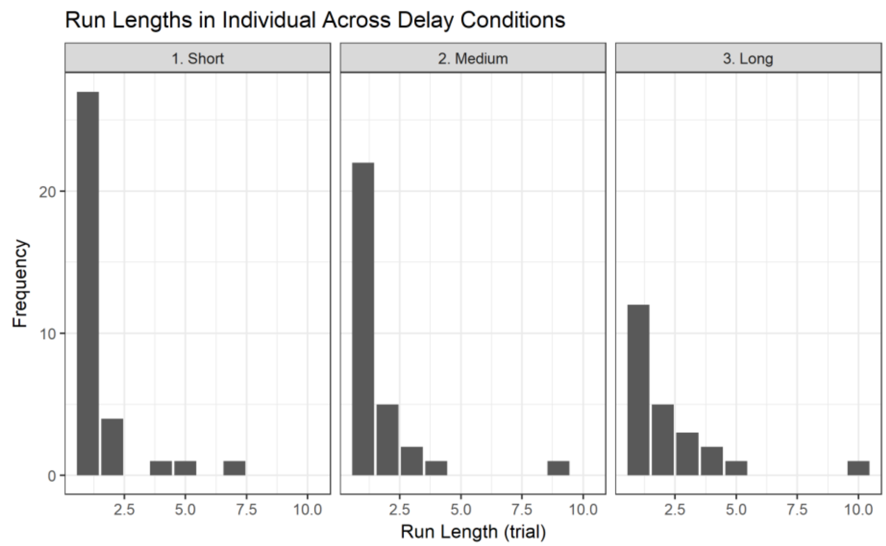
Age distribution of the participants.

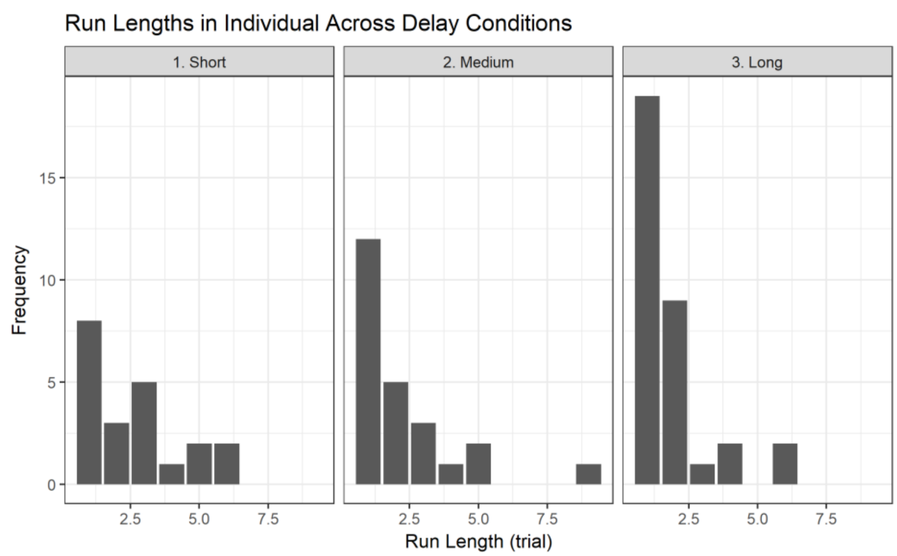
B

Question	Answer options
Age	1. 18-30 2. 31-40 3. 41-50 4. 51-60 5. 61 >
Gender	1. Male 2. Female 3. Other
Are you left or right handed?	1. Right 2. Left
Which hand(s) did you use on the keyboard for this experiment (on average across trials)?	1. Right 2. Left 3. Both
What was your goal?	(Randomized order)

	1. To collect as many apples as possible 2. To gather evenly from both trees 3. To get the same number of apples from each tree 4. Detect the pattern of when a tree of my choice would give an apple 5. Other
Which variables do you think changed throughout the game?	(Randomized order) 1. Waiting time outside or in front of the gate 2. Rate of apples 3. Walking speed of the avatar 4. All options
Which tree gave you more apples based on your experience?	1. Right 2. Left 3. Did not notice a difference
How well do you think you did in this game (according to your goal) on a scale of 1 – 5 where 1 is "Very badly" and 5 is "very well"?	Scale from 1 to 5

C





Run length differences between two individuals. The first one shows a pattern seen in mice – in the short delay the agent mostly switches, in the long delay the decay rate shows that they stick to one option for longer before switching. An opposite pattern is seen in the second individual.