Aging of the Exploring Mind: Older Adults Deviate more from Optimality in Complex Choice Environments

Job J. Schepens
(job.schepens@fu-berlin.de)
Center for Cognitive Neuroscience Berlin
Freie Universität, Berlin 14195, Germany

Ralph Hertwig, Wouter van den Bos
({hertwig,vandenbos}@mpib-berlin.mpg.de)
Center for Adaptive Rationality
Max Planck Institute for Human Development, Berlin 14195, Germany

Abstract

Older adults (OA) need to make many important and difficult decisions about their health, finances, etc. Often, there are too many options available to explore exhaustively, creating the ubiquitous tradeoff between exploration and exploitation. How do OA make these complex tradeoffs? Here, we investigate age-related shifts in solving exploration-exploitation tradeoffs depending on the complexity of the choice environment. OA reliably chose the most rewarding options less often than younger adults (YA) did. Also, choices of OA were less well accommodated by an optimality model. One possible explanation is that OA are less tuned to the uncertainty associated with the options and process this information sub-optimally. This insight may be useful to study how mental representations in complex choice environments can be enhanced, which would help maintaining goal-directed learning later in life.

Keywords: Cognitive aging, learning mechanism, complex choice environment, N-armed bandits, optimality models

Introduction

In today's aging societies, older adults (OA) are increasingly required to make cognitively demanding decisions about their workplace, health, finances, end of life care, etc. These choices are made on a daily and continuous basis (e.g., "How can I make sure that I eat healthy today?") and they often have long term consequences. They require decision makers to keep long-term goals in mind, which may be difficult in cognitively demanding complex choice environments.

Although cognitive decline generally hinders decision-making, the onset and speed of cognitive decline vary widely across individuals (Belsky et al., 2015). Generally, experience-invariant fluid cognitive abilities decline steadily after the second or third decade of life, whereas experience-dependent crystallized abilities peak later (Hartshorne & Germine, 2015). Depending on their age, adults adapt their decision-making strategies to the fluid or crystallized resources available to them (Y. Li, Baldassi, Johnson, & Weber, 2013).

Here, we aim to study possible learning mechanisms that can underlie age-related performance changes. Decision makers generally have access to a number of learning mechanisms (Daw & O'Doherty, 2014), including Pavlovian reflex based learning (Q. J. Huys et al., 2015), habitual experience-based learning (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006), goal-directed learning (Doll, Duncan, Simon, Shohamy, & Daw, 2015). Goal-directed choices are prospective, in the sense that they require planning, and these are thus cognitively demanding. Goal-directed learning can make use of some internal model, so that it can change strategies flexibly and without as much training as experience-based learning. Habitual decisions have been related to a dorsolateral striatal to sensorimotor cortex control loop while goal-directed decisions have been related to a dorsomedial striatal to ventromedial and lateral prefrontal cortex control loop (Daw & O'Doherty, 2014; Doll et al., 2015). Goal-directed learning explicitly makes use of knowledge of relevant task contingencies. In contrast, experience-based learning (similar to model-free reinforcement learning) may not update expectancies for options that require a model of relevant task contingencies.

Goal-directed learning is impaired in OA and this impairment has been associated to lower activation in prefrontal cortex areas (Eppinger & Bruckner, 2015; Eppinger, Walter, Heekeren, & Li, 2013). It has been reported that OA ineffectively rely relatively more often on experience-based strategies and that this maladaptive behavior can arise from (white matter integrity) changes in the ventromedial and lateral prefrontal cortex (Chowdhury et al., 2013; Eppinger et al., 2013; Samanez-Larkin & Knutson, 2015).

Age-related differences in recruitment of goal-directed choices have been elicited using different tasks (Eppinger & Bruckner, 2015; Eppinger et al., 2013; Wit, Vijver, & Ridderinkhof, 2014; Worthy, Cooper, Byrne, Gorlick, & Maddox, 2014). These have generally manipulated task contingencies that affect the reward structure of decisions. Age-related goal-directed strategy differences have been exposed in 2-step sequential choices in which participants track contingencies between actions and delayed rewards (Eppinger et al., 2013). Also, Worthy and colleagues (Worthy et al., 2014) have compared prospectively rewarding decision strategies and immediately rewarding decision strategies. Here, YA favor decisions that are not good now but improve future states whereas OA more often prefer decisions that are good now but worse in the future. Furthermore, age-related goal-

directed strategy differences can account for responses to devaluation of existing goals (Wit et al., 2014). A common pattern across these task manipulations relates to the way exploratory information is taken into account.

The possibility to be able to distinguish between different learning mechanisms based on performance in a certain task may depend on the relative complexity of the task itself. An ecological rationality perspective would predict that task performance crucially depends on the complexity of the choice environment, because of the adaptive fit between the mind to the environment (Mata et al., 2012). For example, when an environment affords a simple strategy, YA and OA adults will also use that strategy instead of more cognitively demanding strategies (Mata, Schooler, & Rieskamp, 2007). Accordingly, simpler choice environments can be studied to find out at what level of task complexity, age-related differences start to disappear.

The present study

In N-armed bandit problems, adults make inferences about the best risky option available to them by sampling information from a choice environment. Rewards are consequential, meaning that all rewards matter for performance. This ensures that participants need to trade-off risky exploratory sampling and exploitation. Participants try to find options that give them the most money while having to minimize sampling from low-reward options. It is unclear what information OA process differently than YA when it comes to continuing, switching, or stopping sampling as the choice environments becomes more complex. We hypothesize that age-related differences can arise in N-armed bandit tasks when OA rely less on goal-directed learning, in particular as the cognitive demands of the task increase. We assume that larger choice environments in N-armed bandit problems are relatively more cognitively demanding because of a larger search space, which increases the amount of information that needs to be processed.

We aim to shed light on age-related differences in participants' representations of expected future reward. Expected reward rises with adequate, but not excessive, exploration. Such "smart" exploration requires one to have a good representation of the task and its structure. Such a representation typically weighs already observed rewards by the degree to which an option has been explored. This would thus involve an internal model of the task contingency between average payoff and average payoff uncertainty. A similar observation was made by Worthy and colleagues (Worthy et al., 2014). If OA over-rely on exploitation to maximize reward in current states, their rewards in future states suffer. Such short-term planning more closely resembles experience-based learning.

We predict a large age effect on exploration as based on a recent aging study that shows that OA explore less than YA do (Frey, Mata, & Hertwig, 2015). There, OA sampled less in a task with a non-consequential sampling phase, depending on the order and complexity of the problems. It is unclear whether cognitive aging exposes similar behavioral differences in exploration that is consequential. If it is similar, we should observe that OA explore less adequately in consequential sampling. Specifically, we expect that OA represent long-term reward inaccurately so that they arrive slower at the higher reward options. We also similarly expect that the expected effects are stronger in more complex choice environments (Frey et al., 2015).

To summarize, we expect that OA achieve relatively lower performance, the more complex the choice environments gets, as adapting to the tradeoff becomes more cognitively demanding. We further expect to see a less adequate exploration phase compared to YA. Next, we describe methods and results from three kinds of data analyses: overall performance, performance over time, and comparisons to optimality. We end with a discussion.

Methods

Participants

32 older adults (OA, M_{age} = 70.5, 52% female) and 29 younger adults (YA, M_{age} = 24.3, 38% female) participated in this study. All participants were healthy, right-handed, native German speakers with normal or corrected to normal vision, and without a history of psychiatric or neurological disorders. Compensation amounted to about 12 Euro per hour, including bonus money that depended on performance. Participants were recruited using advertisements.

Task

The task of the participants was to maximize the sum of rewards in a total of 16 alternating four and eight-armed bandit problems. Rewards could be earned by selecting pictures of casino-style gambling machines presented on a computer screen using a keyboard or mouse. The gambling machines provided random rewards (1 or 0) with a hidden probability that was specific to each machine. The rewards were displayed on the respective bandit after each play and they were then added to the sum of rewards. Participants had 100 trials to explore the hidden reward probabilities and to exploit those machines that give rewards most often. All information necessary to make an optimal choice at any point in time was always present on the computer screen of the participant. This means that each bandit showed the number of plays so far as well as the probability of a reward based on the rewards seen so far. Because of this information, participants do necessarily not have use memory at all to solve the task optimally. Of course, participants still need to figure out how they want to trade off exploration and exploitation. Participants indicated in a post-task questionnaire that they found this information helpful.

Procedure

Participants were instructed 1) to maximize the sum of rewards, 2) how the task looked and worked, 3) that each trial is independently generated, and 4) that the best gambling machine for all 16 problems has $p_{opt} = .6$. The last instruction was used to help minimize a learning effect over the course

of the experiment. All participants had taken part in an unrelated fMRI study on risk-taking preference several weeks beforehand. Ethics approval was granted by the Institutional Review Board of the Max Planck Institute for Human Development.

Design

The experiment made use of a repeated within-subject condition (four vs eight options), and a between-subject condition (age group). We designed the options such that the game remained challenging over the 100 trials, meaning that the options will not always be distinguishable within one problem, even for an optimality model. All participants saw the same randomly generated rewards for all 16 problems: i.e. rewards were randomly generated once beforehand and then read from the same file for each participant. This allowed comparison of the problem difficulty across participant groups as well as a reduction of an unnecessary source for variance in performance while keeping random aspect of the task intact. Four different problem orders were generated and counterbalanced across participants. Two different orders started with four options and two different orders started with eight options. Between problems, performance was displayed on the screen and a keypress was required to continue with the next problem. Participants in both groups took about half an hour to finish the experiment. The minimum response time was set to 200 milliseconds.

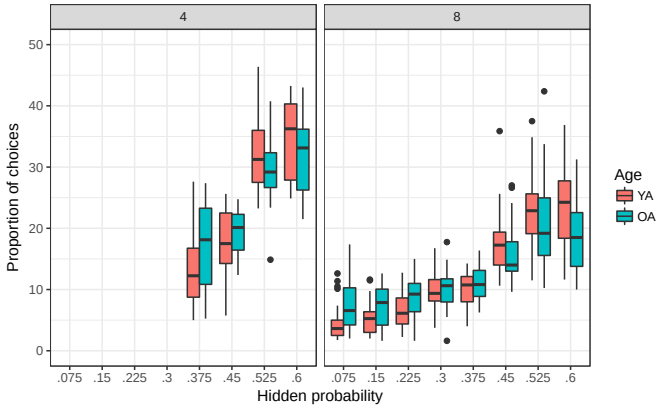

Figure 1: Variations in the choice proportions for every option for both age groups and complexity conditions.

Results

We first investigated age-related differences in task performance. Proportions of choices for each option revealed that YA sampled reliably more often from the better options, see Figure 1. OA chose the highest-expected-value option about 5% less often than YA did in both four and eight option environments (four option 95% HDI: .003 - .093; eight option 95% HDI: .018 - .096; Bayesian ANOVA with logit function and broad prior). The task thus seems to be more difficult for OA as based on their behavior.

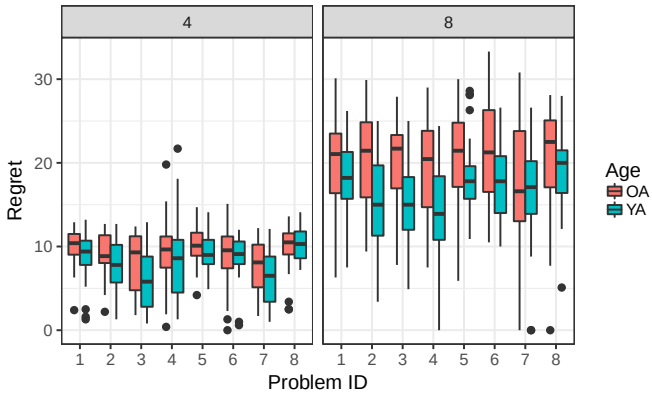

Figure 2: Variation in performance across the 16 different problems in the task.

The difficulty of the task can be specified further into the relative difficulty of all of the 16 distinct bandit problems. However, we need to introduce another measure of task performance to be able to compare performance across all of the options and not only for the trials where the best option is chosen. We choose to measure regret as it generalizes over the specific outcomes of the random number generation process. In this experiment, regret can be computed as:

$$R_T = \sum_{i=1}^{100} \left(p_{opt} - p_{B(i)} \right) \tag{1}$$

where $p_{B(i)}$ is the hidden probability of a reward for the chosen bandit. It follows that randomly behaving agents get a total regret of 11.25 points in a four option environment and 26.25 points in an eight option environment. Overall, the age effect on regret was large (Cohen's d .766) in the eight option condition (M_{OA} = 19.87, SE = 1.43, M_{YA} = 16.84, SE 1.36) and medium to small (Cohen's d .364) in the four option condition (M_{OA} = 9.16, SE = .9, M_{YA} = 8.17, SE 1.03). Using this new measure, we can also show that age-related differences depend not only on the complexity of the choice environment in terms of the number of choice options, but they also depend on the relative difficulty in terms of the intricacies of the random number generation process itself, see Figure 2. For some of the decision problems, the random number generator is favorable to OA so that it even completely obscures age-related differences. Furthermore, it is clear that the degree of individual differences scales with the complexity of the choice environment. This does not necessarily have to be surprising in the light of the exponential increase in the number of possible choice profiles in a larger search space as well as the fact that the eight option conditions contains relatively worse options (e.g. .075 differs more from .6 than .375). Note that some bar plot whiskers even reach the 0 regret line, indicating that there were constant choice profiles that contain the best option all the way from the beginning.

Second, we investigated how performance differences

emerged using the shapes of the exploration-exploitation trade-offs over trials within the choice profiles. We first looked at regret as a measure of overall performance over trials. There, we observed reliable age-related differences for both conditions. For both groups, average performance was reliably better compared to a random agent and reliably worse compared to an optimal agent from very early on (<20 trials), see Figure 3.

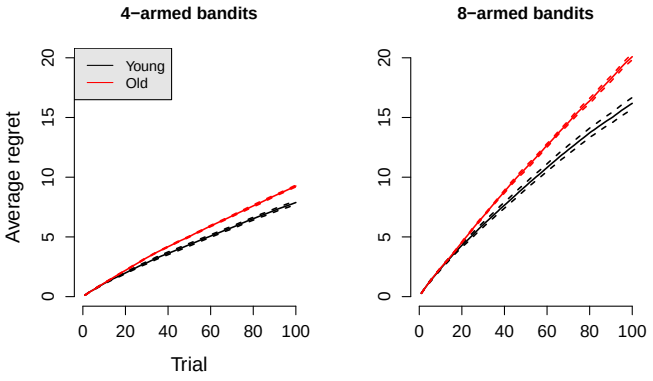

Figure 3: Regret at every trial for both groups with standard errors displayed as dashed lines

We further analyzed the choice trajectories specifically for every choice option to compare age-differences for each of the options specifically over time, see Figures 4 and 5. For every trial, the values of the solid lines in these trajectories represent the average number of times that participants chose an option. After about 25 trials, the 3rd best option seems to stop overlapping reliably with the 2nd best option. This was confirmed by comparing standard error bounds. For OA, the same separation exists between the second and third option, but the separation only becomes reliable after twice as many trials, see the right panel of Figure 4. The trajectories further reveal that the task is relatively difficult as the trajectories stabilize only marginally. Furthermore, the deviations within each of the solid lines indicate individual differences again. Even within each age-group, the participants used widely different strategies. The eight options setting shows a pattern that is comparable to the four option condition. YA separate reliably between the better three options and the worse five options after about 50 trials. OA do this after about 75 trials. Together, the choice trajectories show that already from the beginning onwards, YA choose more often from the better options. This may indicate that the seeds for a performance difference later on are already sown early on.

Third, we wanted to know how far people's behavior deviates from optimality. Therefore, we compared the shapes of the mean observed exploration-exploitation trade-off trajectories over trials to shapes as predicted from an optimality model that was fitted to the participant's choices using a softmax rule. For the optimality model, we chose Thompson sampling (Thompson, 1933), but we observed that dif-

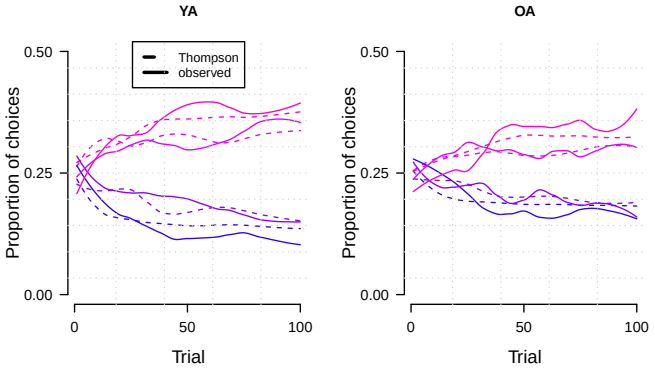

Figure 4: Observed and predicted choice proportions at every trial for the four option environment. The darkness of the lines corresponds to the value of the hidden probabilities, where lighter colors represent higher probabilities. Local polynomial regression was used as a moving window to smooth the trajectories (using a neighborhood of 40% of all points where neighboring points are also being weighted by their distance tricubically).

ferences in regret for similar algorithms are small in the context of the present task. Thompson sampling uses an inverse cumulative distribution function (also known as percentage point function or quantile function) that is used to choose the bandit with the highest certainty that the hidden probability of a bandit is smaller or equal than some randomly generated value. This way, the model minimizes uncertainty that there exists a better option by making sure that the probability of choosing a certain bandit is proportional to the probability of it being the best bandit. By taking uncertainty into account, the strategy affords a way of rapidly changing its current strategy if not only the mean of a certain bandit gets overtaken by another mean, but the whole posterior probability distribution. Conceptually, the algorithm keeps track of beliefs about the hidden probabilities of the bandits and then updates these beliefs each time after seeing an outcome. As shown in Table 1, the algorithm is initialized by setting a uniform prior for all options. The algorithm then plays option x proportional to the probability of it being the best. Finally, it updates its priors using the newly collected information. Thompson sampling achieved an expected regret of 6.5 points in the four options environment and a regret of 11.0 points in the eight options environment. 5 out of 32 OA and 9 out of 29 YA achieved performance within 10% of this performance level in the four option environment, while this was 1 and 4 respectively in the eight option environment.

In order to fit Thompson sampling to participant's choices, some way of scaling the predicted choices to decision probabilities is necessary. Otherwise, if a participant selects an option that the model would never select, the minus log likelihood of the model will be infinitely high. We made use of a standard softmax rule with an inverse temperature pa-

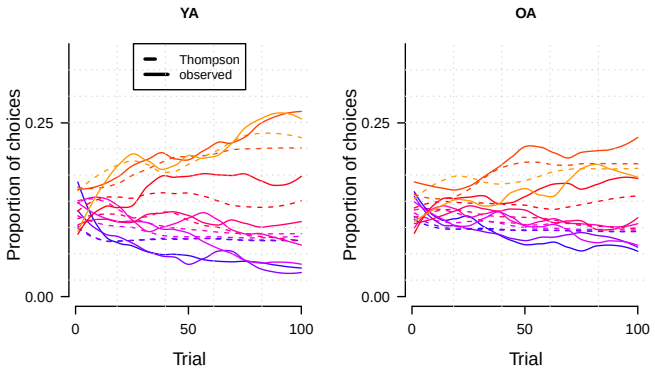

Figure 5: Observed and predicted choice proportions at every trial for the eight option environment, see also Figure 4

rameter θ ranging from 0.003 to 30. θ controls how choices match to prediction, with higher values resulting in more random behavior. We first fitted Thompson sampling to every problem separately (using both fmincon in Matlab and optim in R). We then computed the median deviance from Thompson sampling for every participant and plotted those values in Figure 6. The distributions of goodness of fit differed across groups. OA deviated more from Thompson sampling than YA did. The distributions are naturally cut off at the points of random choice, which are $100 \times \log_2 25\% = 138.6$ for four options, and $100 \times \log_2 12.5\% = 207.9$ for eight options. The deviances for the YA are slightly more dispersed and relatively more OA reach a point of deviation from optimality that equals to random behavior. Furthermore, the median θ was lower in YA for both conditions, further indicating less randomness in YA than in OA. For four options, the median θ for YA was .67 (IQR .75) and for OA 1.11 (IQR 4.06), and for eight options the median θ for YA was .52 (IQR .36) and for OA it was .87 (IQR 1.61). The median θ was also generally lower in the eight option environment, indicating that the model predictions match participants' choices better. However, this might actually simply be a consequence of the task structure. Four relatively low-probability options were added in the eight option condition so that the number of competing options stays the same while changing proportionally to the total number of options. Due to the within-subject design, we could also analyze individual changes for both age groups across conditions. OA more than YA decreased their median θ in the eight option condition compared to the four option condition, with a median decrease in θ for YA at .16 (IQR .37) and for OA at .34 (IQR .95). This larger decrease may indicate more of a strategy change depending on the choice environment. The variation in these decreases was also larger for OA. OA adults are thus a less homogeneous group than YA, possibly indicating more strategy changes.

We then compared the optimality model's predictions to observed choice profiles. These comparisons are plotted in Figures 4 and 5 using dashed lines. We used the median θ over all eight problems within each condition as a participant-specific data scaling parameter. Predictions from a fitted Thompson sampling model resulted in remarkably accurate trajectories for both groups and for both conditions. The overlapping trajectories may be useful to investigate under- and oversampling from the respective options although this is tentative due to large confidence bounds around each of the respective trajectories. We observed that the order of the dashed lines is more accurate than the orders of the solid lines in the beginning of all the four graphs. This seems to indicate rapid exploration in Thompson sampling and that both YA and OA explore less rapidly, with OA taking the longest.

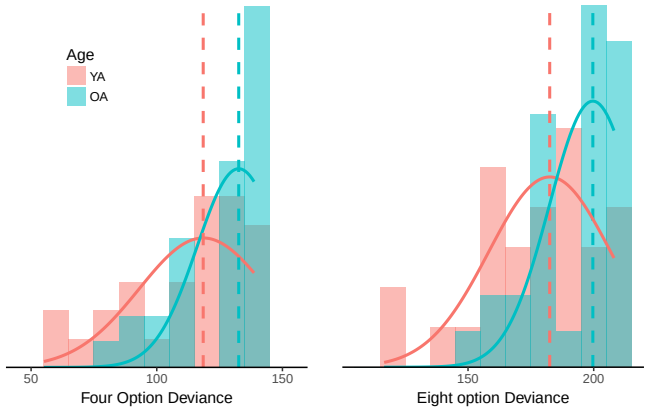

Figure 6: Median deviance from Thompson sampling for every individual with a derived normal curve (solid line) and mean (dashed line)

Discussion and Conclusions

This study aimed to identify changes in the ways OA vs. YA make goal-directed choices depending on the complexity of the choice environment. We found a large age-related effect on performance in a typical eight-armed bandit task and a smaller effect in a four-armed bandit task. Choice trajectories showed that age effects are already observable in the very early exploration stage (<20 trials). Reliable differences between individual choice options became apparent after about 25 to 50 trials, depending on the size of the search space as defined by the number of options in the choice environments. YA also deviated less from optimality than OA did, especially in the early stages of the choice profiles. These early stages require fast exploration using not only average rewards but also their associated uncertainty. This was illustrated here using Thompson sampling, which is a kind of randomized probability matching algorithm. Participants diversified their choices in a remarkably similar way to Thompson sampling, in line with previous work (Konstantinidis, Ashby, & Gonzalez, 2015; Speekenbrink & Konstantinidis, 2015). Furthermore, there were age-related differences and variations in the inverse temperature parameter estimates across choice environments, indicating that OA might rely on simpler learning mechanisms that take less of the important information into

account, in particular in the more complex environments.

Future work may use the present comparison of behavior to optimality to navigate the space of cognitive models. Specifically, the exploration phase seems to happen sub-optimally in OA and in a more varied way. Favoring short-term rewards could be a sign of a learning mechanisms that sub-optimally represent long-term rewards. Performance in the present task benefits from a weighting scheme between rewards and uncertainty. Representation of future reward suffers from a failure to capture such a scheme, possibly reflecting the use of less goal-directed and more experience-based learning. A more varied and reduced cognitive processing of necessary information such as uncertainty would be able to successfully account for the observed age-related changes in behavior.

To conclude, it seems that OA rely longer on simpler learning strategies, the more demanding the choice environment becomes. Identifying these effects is typically neglected in neuroeconomic models of decision making. In the context of cognitive aging this is particularly important for empowering aging decision makers to navigate cognitively demanding choice environments.

Table 1: Thompson sampling

Step	Computation
Init	$beta(1,1)$
Choose	$softmax(q,\theta)$
	$q = max(qbeta(p,\alpha,\beta))$
Update	$\alpha = 1 + wins$
	$\beta = 1 + pulls - wins$

Acknowledgments

European Commission Horizon 2020 MSCA No 708507.

Open Practices

Elicitation materials, all data, and analysis scripts can be found on `https://github.com/jobschepens/Bandits/`

References

Belsky, D. W., Caspi, A., Houts, R., Cohen, H. J., Corcoran, D. L., Danese, A., ... Moffitt, T. E. (2015). Quantification of biological aging in young adults. *Proceedings of the National Academy of Sciences*, *112*, E4104–E4110.

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Dzel, E., & Dolan, R. J. (2013). Dopamine restores reward prediction errors in old age. *Nature Neuroscience*, *16*, 648–653.

Daw, N. D., & O'Doherty, J. P. (2014). Chapter 21 - Multiple Systems for Value Learning. In P. W. G. Fehr (Ed.), *Neuroeconomics (Second Edition)*. San Diego: Academic Press.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876–879.

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, *18*, 767–772.

Eppinger, B., & Bruckner, R. (2015). Chapter 4 - Towards a Mechanistic Understanding of Age-Related Changes in Learning and Decision Making: A Neuro-Computational Approach. In T. M. Hess & J. S. E. Lckenhoff (Eds.), *Aging and Decision Making*. San Diego: Academic Press.

Eppinger, B., Walter, M., Heekeren, H. R., & Li, S.-C. (2013). Of goals and habits: Age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience*, *7*.

Frey, R., Mata, R., & Hertwig, R. (2015). The role of cognitive abilities in decisions from experience: Age differences emerge as a function of choice set size. *Cognition*, *142*, 60–80.

Hartshorne, J. K., & Germine, L. T. (2015). When Does Cognitive Functioning Peak? The Asynchronous Rise and Fall of Different Cognitive Abilities Across the Life Span. *Psychological Science*, *26*, 433–443.

Huys, Q. J., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., ... Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, *112*, 3098–3103.

Konstantinidis, E., Ashby, N. J., & Gonzalez, C. (2015). Exploring Complexity in Decisions from Experience: Same Minds, Same Strategy. In *37th annual meeting of the Cognitive Science Society (CogSci 2015)* (pp. 23–25).

Li, Y., Baldassi, M., Johnson, E. J., & Weber, E. U. (2013). Complementary cognitive capabilities, economic decision making, and aging. *Psychology and Aging*, *28*, 595–613.

Mata, R., Pachur, T., Von Helversen, B., Hertwig, R., Rieskamp, J., & Schooler, L. (2012). Ecological rationality: A framework for understanding and aiding the aging decision maker. *Decision Neuroscience*, *6*, 19.

Mata, R., Schooler, L. J., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging*, *22*, 796–810.

Samanez-Larkin, G. R., & Knutson, B. (2015). Decision making in the ageing brain: Changes in affective and motivational circuits. *Nature Reviews Neuroscience*. *16*, 278–289.

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and Exploration in a Restless Bandit Problem. *Topics in Cognitive Science*, *7*, 351–367.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 285–294.

Wit, S. de, Vijver, I. van de, & Ridderinkhof, K. R. (2014). Impaired acquisition of goal-directed action in healthy aging. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 647–658.

Worthy, D. A., Cooper, J. A., Byrne, K. A., Gorlick, M. A., & Maddox, W. T. (2014). State-based versus reward-based motivation in younger and older adults. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 1208–1220.