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. 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? We investigated age-related shifts in solving exploration-exploitation tradeoffs depending on the complexity of the choice environment. Participants played four and eight option bandit problems with numbers of gambles and average rewards available on the screen. OA reliably performed worse in a more complex choice environment and were also more deviant from an optimality model (Thompson sampling), which keeps track of uncertainty beyond just the mean or last reward. OA seem to process important information in more complex choice environments sub-optimally, suggesting limited representations of future rewards. This interpretation fits to multiple contexts in the complex cognitive aging literature, in particular to the context of challenges in the maintenance of goal-directed learning.

Introduction

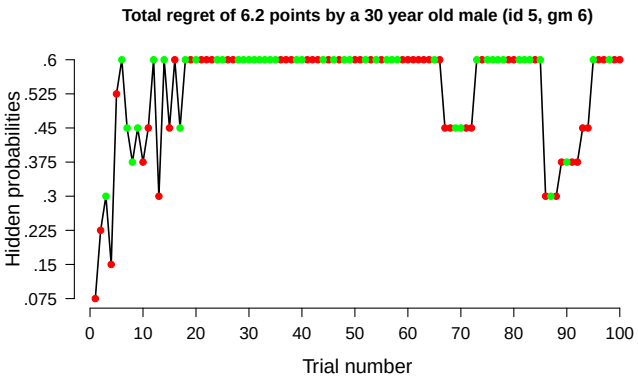
In today's aging societies, more and more older adults (OA) are making cognitively demanding decisions about work, finances, their health, etc. Many such decisions benefit from thinking about future goals because the options available create explore-exploit tradeoffs. How do OA usually respond to these cognitive challenges in increasingly complex choice environments?

Decision makers generally have access to a number of learning mechanisms, habitual experience-based learning, and goal-directed learning. Goal-directed learning depends on some internal model, so that learning can be adapted flexibly, for example like when managing a research project. Habitual learning has been related to a dorsolateral striatal to sensorimotor cortex control loop while goal-directed learning has been related to a dorsomedial striatal to ventromedial and lateral prefrontal cortex control loop (Daw & O'Doherty, 2014). Importantly, 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). OA rely relatively more often on experience-based learning, which may 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).

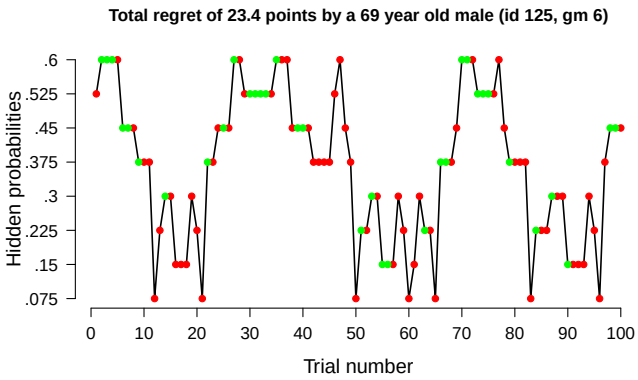
It is unclear how such changes in learning mechanisms in OA depend on the relative complexity of a task. Such a dependency would be likely, however, from the perspective of ecological rationality (Mata et al., 2012), which focusses on adaptation effects between the mind and the environment. For example, when OA need to explore among many options in order to choose between them later, OA rely on more minimal exploration strategies than YA (Frey, Mata, & Hertwig, 2015). Here, we study such age-related performance changes in explore-exploit tradeoffs with varying cognitive demands. Analyzing effects of the complexity of choice environments this way could help to better understand the effects of task demands on age-related changes in learning mechanisms.

We used typical N-armed bandit problems to study changes in learning mechanisms across choice environments. Participants made inferences about risky options by sampling information from a four and eight option choice environment. Rewards were consequential, ensuring that participants needed to trade-off exploration and exploitation. Participants had to find options that give them the most money while having to minimize sampling from low reward options. N-armed bandit problems are well studied and afford detailed analysis of information processing in terms of continuation and switching behavior. Theoretically, expectations of future reward should rise with adequate, but not excessive, exploration. Such "smart" exploration requires one to have a good representation of the task and its structure, which 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 (see also Worthy, Cooper, Byrne, Gorlick, & Maddox, 2014). Good performance in the task thus depends on learning mechanisms that use adequate future reward representations while performance anomalies will involve inadequate future reward representations.

We hypothesized that OA achieve lower performance and arrive slower at the higher reward options, depending on the number of options. If OA focus more on reward in current states, their rewards in future states should suffer. Such short-term planning would more closely resemble experience-based learning rather than goal-directed learning. OA not arriving



(a) A more effective strategy that seems to take into account uncertainty.



(b) A less effective strategy that seems to depend almost only on the outcome of the last trial.

Figure 1: Example eight-option choice profiles. Green indicates rewards and red indicates no rewards on a trial.

at the higher reward options at all would show a lack of an explore-exploit trade-off. We assumed that problems with a larger number of options are relatively more cognitively demanding because of a larger search space, which increases the amount of necessary information processing and representation.

Next, we describe methods and results from six kinds of data analyses: choice proportions statistics, choice proportions over trials, regret over trials, comparisons to an optimality model, comparisons to a fitted optimality model, and one-step ahead predictions of a fitted optimality model. We end with a discussion.

Methods

Participants 32 older adults (OA, M_{age} = 70.5, 65-74, 38% female) and 29 younger adults (YA, M_{age} = 24.3, 19-30, 45% 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. There were no group differences in gender proportion, educational level, and socioeconomic status. Compensation amounted to about 10 Euro per hour, plus on average 2 Euro performance-dependent bonus. 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. Participants had 100 trials for every problem to explore the hidden reward probabilities and to exploit those machines that give rewards most often. Remaining trials were displayed on the screen. Also, every bandit showed the number of plays so far and the probability of a reward based on the observed rewards so far. This information is sufficient to make an optimal choice at any point in time, reducing the role of working memory. Of course, participants still need to figure out how they want to trade off exploration and exploitation. 89% of YA and 70% of OA (p = .14, test of equal proportions) indicated in a post-task questionnaire that "the extra information regarding the options" was 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 in every individual problem had $p_{opt} = .6$ (to help comparability across problems). 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 chose the other hidden probabilities in steps of .075 below .6. Reliably finding the better options thus required a significant part of exploration out of the 100 available trials. See also Figure 1 for example choice profiles and the unique hidden probabilities. All participants saw the same randomly generated rewards for all 16 problems. 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 the probabilistic character 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

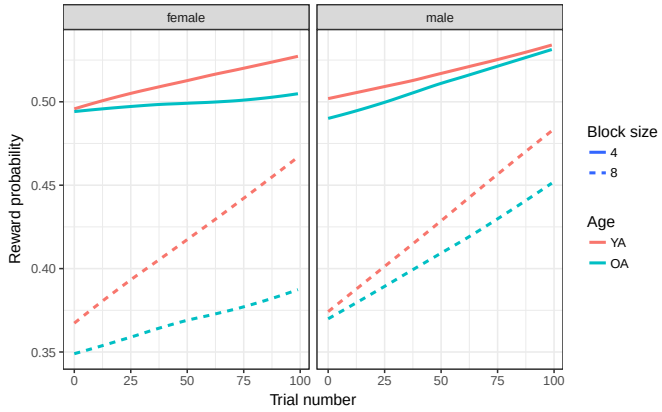


Figure 2: Predictions from a mixed effects model with the hidden reward probabilities of participant's choices as dependent variable and interaction effects between age, number of options, trial, and gender.

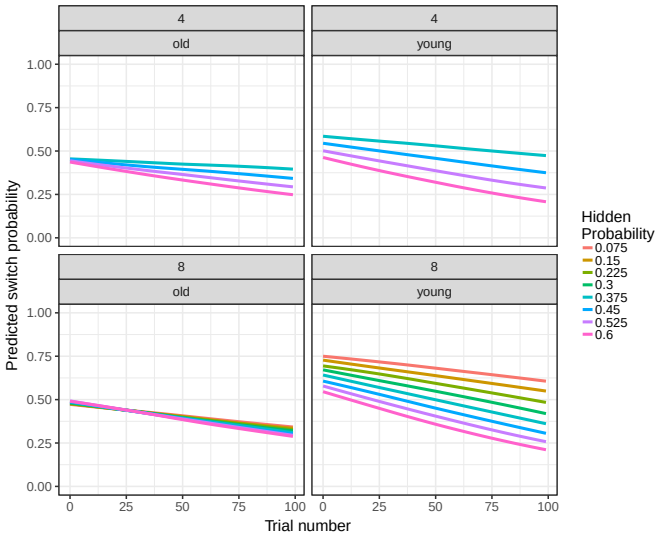


Figure 3: Predictions from a mixed effects model with switching as dependent variable and three-way interaction effects between age, number of options, trial, reward, and the hidden probabilities of participant's choices.

finish the experiment. The minimum response time was set to 200 milliseconds.

Results

We first investigated age-related differences in task performance. Proportions of choices for each option revealed that OA chose the option with the highest hidden probability about 5% less often than YA did in both four and eight option conditions (four option 95% HDI: .003 - .093; eight option 95% HDI: .018 - .096; Bayesian ANOVA with logit function and broad prior), see Figure 4 for the differences for all options. We also tested a linear mixed effects model with the hidden probability of every chosen bandit as dependent variable and with participant ID, problem ID, and bandit position ID as random effects. We used Satterthwaite's approximations of p-values (*** indicating $p < .001$). We found negative interaction effects for OA in eight options (B = -.021***) and for OA in eight options over trials (B = -.011***). Together, these indicated a lower performance for OA in eight options, as well as an increasingly lower performance over trials. We also found a positive interaction effect for both YA and OA in eight options over trials (B = .021***), as participants could improve relatively more over time for eight options. There still was a main negative effect of eight options (B = -.092***) and a main positive effect of trial number (B = -.009***). No significant difference or decrease over trials for OA remained, so the age effect is captured only by the higher-level interactions. Furthermore, we also controlled for gender effects, which indicated that male OA performed better over trials (B = .010***) and that male OA performed better with 8 options (B = .024***), and that males generally performed better (B = .003**). For visualizing these high level interactions, we generated predictions from this model using the package merTools, see Figure 2. Note that the visualization does not show raw data and that the differences in intercepts and slopes for the lines displayed should be interpreted in the light of all data included in the model. Besides performance, we used a similar statistical analysis to test age differences in switching probability over time. The resulting logistic regression model included age, number of options, trial number, the hidden probability, the reward for the participant's choices, and all three-way interactions. Together, the estimated effects on switching indicated that OA switch less often (**), OA switch away less often after sampling from an option with a relativey low hidden probability (***), especially in eight options over trials (**). Beta's were not easily comparable for this model. We again generated predictions from this model to visualize these switching patterns, see Figure 3.

Second, we examined development of age-differences in choice proportions over trials, see Figures 6a and 6b. For every trial, the solid lines represent the average number of times that participants chose an option. The local instabilities in the trajectories may result from individual differences and variation across the several problems. On average, the third best option stops overlapping with the second best option after about 25 trials for YA. For OA, the same separation exists between the second and third option after twice as many trials, see the right panel of Figure 6a. In the eight option condition, YA separate between the better three options and the worse five options after about 50 trials. OA do this after about 75 trials. These 2-2 and 3-5 separations could reflect the participants' psychologically most salient explore-exploit representations. Together, the choice trajectories show that already from the beginning onwards, YA choose more often from the better options.

Third, we analyzed another measure of performance to compare performance across all of the options at once. We

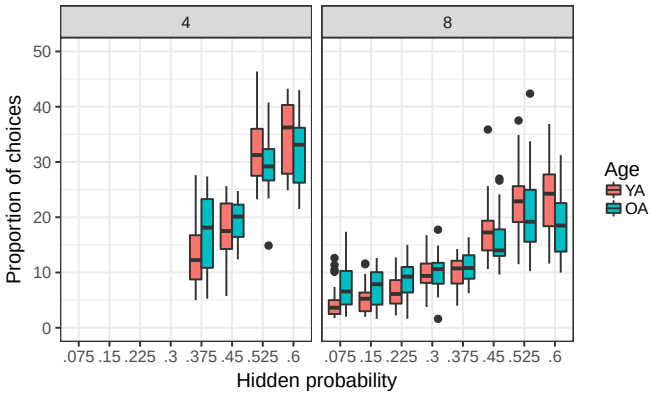


Figure 4: Boxplots of variation in average choice proportion across participants for both choice environments.

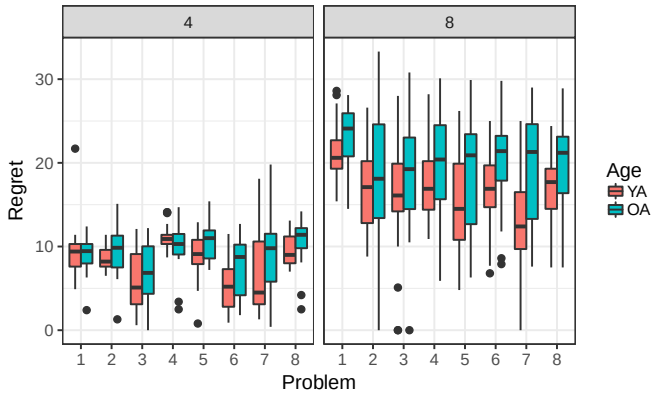


Figure 5: Variation in performance across the 16 different problems in the task.

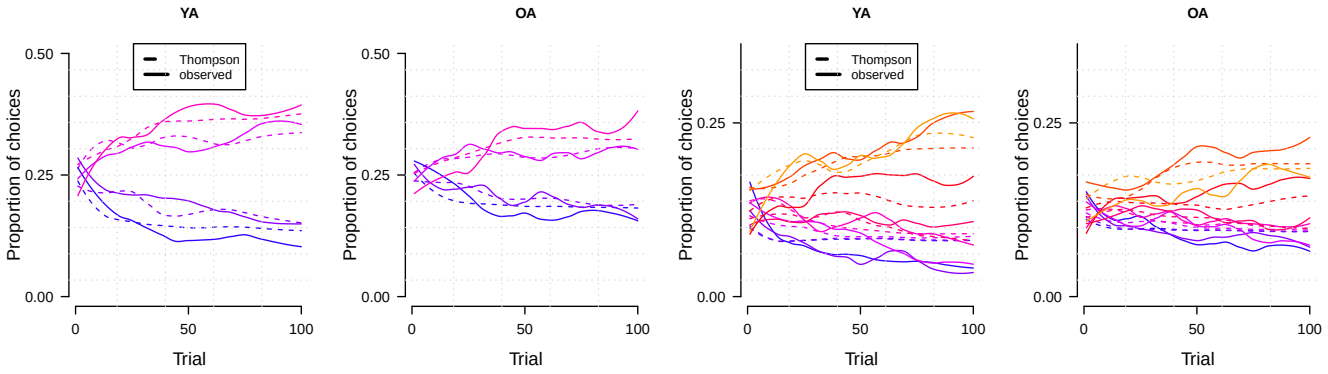
choose to measure regret (a common measure in machine learning) as it generalizes over the specific outcomes of the random number generation process. Regret can be computed as $R_T = \sum_{i=1}^{100} \left(p_{opt} - p_{B(i)}\right)$, where $p_{opt} = .6$ and $p_{B(i)}$ is the hidden probability of a reward for the chosen bandit. It follows that randomly behaving agents get a total regret of 11.25 points for four options and 26.25 points for eight options. Overall, the age effect on regret was large ($p < .01$, Cohen's d .707) for eight options (M_{OA} = 19.87, SE = .79, M_{YA} = 16.84, SE .75) and medium ($p < .05$, Cohen's d .550) for four options (M_{OA} = 9.16, SE = .32, M_{YA} = 8.17, SE .33). These age-related differences varied slightly across the unique problems, which only differed by random number generation, see Figure 5. We also investigated how regret differences emerged using the shapes of the exploration-exploitation trade-offs over trials within the choice profiles. We observed a slowing increase in regret over time in general but increasing age-related differences for both conditions, see Figure 7. Age-differences became significant after trial 24 in eight options and 23 trials in four options. It seems that exploration in OA happens less effectively. Regret was significantly ($p < .05$, t.test) better compared to a random agent (four options: YA after trial 17, OA 32; eight options: YA after trial 15, OA 16).

Fourth, we wanted to know how participant performance differed from optimality. We used Thompson sampling as an optimality model (Thompson, 1933), but we observed that differences 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 algorithm 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 algorithm affords a way of more rapidly adapting its decision 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. 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. See also Table 1. Regret as computed from applying Thompson sampling 29 times to the same games as participants played was significantly worse compared to participants (four options: YA after trial 14, OA 11; eight options: YA after trial 16, OA 16). Expected regret was 6.5 points for four options and 11.0 points for eight options, which is considerably better than YA performed on average (170% larger than the gap between YA and OA for four options and 193% for eight options). Interestingly, 5 out of 32 OA (16%) and 9 out of 29 YA (31%) achieved a median regret score within 10% of Thompson sampling for four options, while this was 1 (3%) and 4 (14%) for eight options. Some individuals were thus able to achieve regret scores similar to Thompson sampling.

Table 1: Thompson sampling in r pseudocode, with n being the number of bandits, x a randomly generated probability, and qbeta for looking up quantiles from the Beta distribution.

Step	Computation
Init	$wins = rep(0,n)$
	$pulls = rep(0,n)$
Choose	$softmax(q,\theta)$
	$q = max(qbeta(x,\alpha,\beta))$
Update	$wins = wins + reward$
	$pulls = pulls + 1$
	$\alpha = 1 + wins$
	$\beta = 1 + pulls - wins$

Fifth, we wanted to know how well a fitted optimality model predicted participant's decisions. Thompson sampling



(a) Four options

(b) Eight options

Figure 6: Observed choice proportions and one-step ahead predictions for fitted Thompson sampling. The color of the lines corresponds to the value of the hidden probabilities, where blue colors represent lower 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)

was fitted to individual games by scaling predicted choices using a softmax function with a fitted inverse temperature parameter θ, ranging from 0.003 to 30 (higher θ values produced more randomness). OA deviated more from fitted Thompson sampling than YA did ($p < .05$, Wilcoxon tests, Cohen's d = .57 for four options and Cohen's d = .53 for eight options). θ was also significantly lower for YA than for OA ($p < .05$, Wilcoxon test, Cohen's d = .96 for four and 2.68 for eight options), indicating more randomness and worse matches to predictions of Thompson sampling in OA than in YA. OA and YA both significantly decreased their median θ for eight options compared to the four option condition ($p < .05$, Wilcoxon tests), see Figure 8. θ for OA significantly varied more in both conditions than for YA ($p < .01$ for both conditions, Wilcoxon tests) and average variation across games was significantly lower in four options for YA, but for OA this was similar in both conditions ($p < .01$ vs. $p = .4$, Wilcoxon tests). In all, OA adults were more random and less homogeneous, possibly indicating more strategy changes.

Finally, we compared the shapes of the mean observed exploration-exploitation trade-off trajectories to shapes from one-step-ahead predictions across all trials. These predictions are plotted in Figures 6a and 6b using dashed lines. We used the median θ of every participant for both conditions as data scaling parameter. The predictions from this fitted Thompson sampling model resulted in accurately ordered trajectories for both groups and for both conditions: The orderings of solid and dashed lines were identical for all four graphs for most trials, except in the first few trials. The latter may indicate more rapid exploration in Thompson sampling and that both YA and OA explore less rapidly, with OA taking the longest.

Discussion and Conclusions

We aimed to identify changes in the ways OA and 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. YA also deviated less from optimality than OA did. Choice trajectories showed that age effects were already observable in the early exploration stage, suggesting that OA explore longer or less efficiently. Theoretically, the 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 similar to Thompson sampling, in line with previous work (Konstantinidis, Ashby, & Gonzalez, 2015; Speekenbrink & Konstantinidis, 2015). Furthermore, OA had higher and more variable inverse temperature parameter estimates across choice environments, indicating more randomness in OA. OA thus rely on less effective learning strategies that consider important information less effectively, in particular in the more complex environments.

Why would OA fail to represent important information like uncertainty or a specific task model? The role of working memory influences should be minimal as this is not strictly necessary to perform well in the task. General "slowing", gender effects, and more cautious risk taking, all of which could favor exploitation of short-term rewards, also mark cognitive aging. We did indeed observe gender interactions, age-differences in reaction times, and in standard neuropsychological test results (working memory, fluid intelligence, and risk-taking). However, as performance is mainly determined by a cognitively costly explore-exploit tradeoff and adequate future reward representations, our findings specifically point towards underreliance on goal-directed learning.

A logical next step is to assess if fits of simple learning strategies can indeed better accommodate OA. Specifically, the exploration phase seems to happen sub-optimally in OA

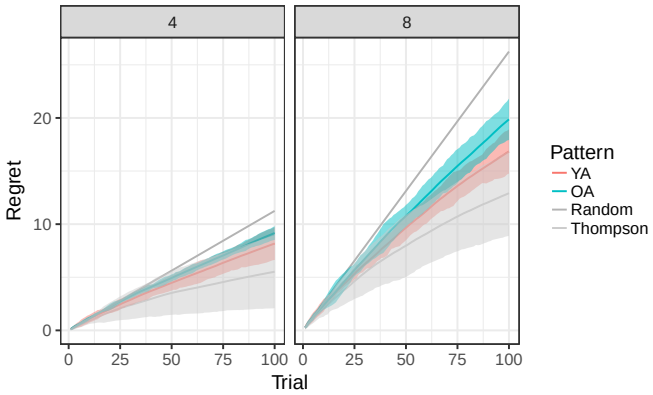


Figure 7: Age-differences in the increase in regret at every trial with standard deviations displayed around the means (standard errors were too narrow to visualize). Regret increased quickly first and increased slower later on, but slower for YA.

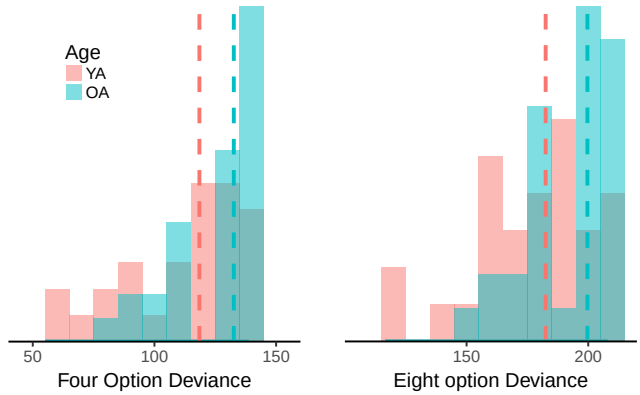


Figure 8: Histogram and medians of median deviance per individual across games from fitted Thompson sampling.

and in a more varied way. Favoring short-term rewards could be a sign of a learning mechanism that sub-optimally represents future rewards. More varied or reduced processing of important information such as uncertainty would be able to account successfully for the observed age-related changes. Furthermore, if the task indeed probes OA to rely less on goal-directed learning, we may also expect differences in connectivity to prefrontal regions (pending analyses). In all, OA may be using less effective learning strategies the more demanding the choice environment becomes. Identifying such task-dependent differences is typically neglected in neuro-computational models of decision-making. In the context of cognitive aging, this may be useful for empowering aging decision makers to navigate cognitively demanding choice environments.

Acknowledgments

JS is supported by European Commission Horizon 2020 MSCA No 708507. Robert Lorenz, Jann Wscher, Dania Esch, and Lukas Nagel supported with participant recruitment, organization, and data collection. Materials, data, and analysis can be found on `https://github.com/jobschepens/Bandits/`.

References

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). Multiple Systems for Value Learning. In P. W. G. Fehr (Ed.), *Neuroeconomics (Second Edition)* (pp. 393–410). San Diego: Academic Press.

Eppinger, B., & Bruckner, R. (2015). 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* (pp. 61–77). 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*.

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).

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(5)*, 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.

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.