Aging of the exploring mind: Older adults deviate more from optimality in complex environments

Job J. Schepens[1], Ralph Hertwig[2], Wouter van den Bos[2]
job.schepens@fu-berlin.de, hertwig@mpib-berlin.mpg.de, vandenbos@mpib-berlin.mpg.de
[1] Center for Cognitive Neuroscience Berlin, Freie Universität Berlin, Berlin, Germany
[2] Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

Introduction

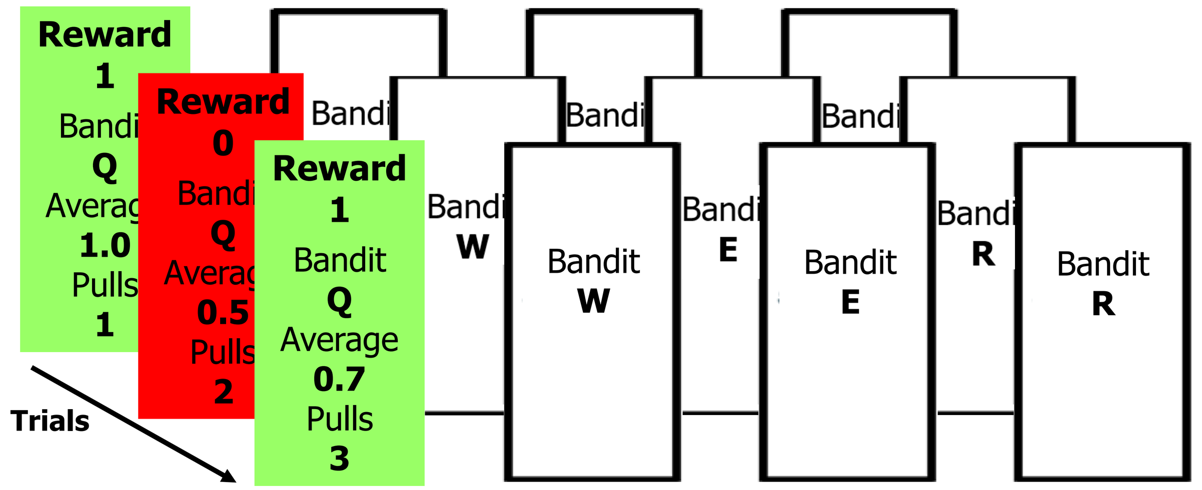
- Exploration-exploitation tradeoffs are ubiquitous in the daily environments of younger adults (YA) and older adults (OA)
  - There are often too many options available to explore exhaustively
  - E.g. managing work activities, health, finances, etc.
- Optimal models such as Thompson sampling achieve high performance by weighing observed rewards with the number of times that an option has been sampled (uncertainty)
- Past work shows how OA under-rely on uncertainty processing, goal-directed learning, and model-based strategies in related tasks (Eppinger, Walter, Heekeren, & Li, 2013; Nassar, Bruckner, Gold, Li, Heekeren, & Eppinger, 2016; Worthy, Cooper, Byrne, Gorlick, & Maddox)
- We study how these mechanisms depend on task complexity, using 4- and 8-armed bandit problems with observed rewards and number of pulls always displayed on the screen

Predictions

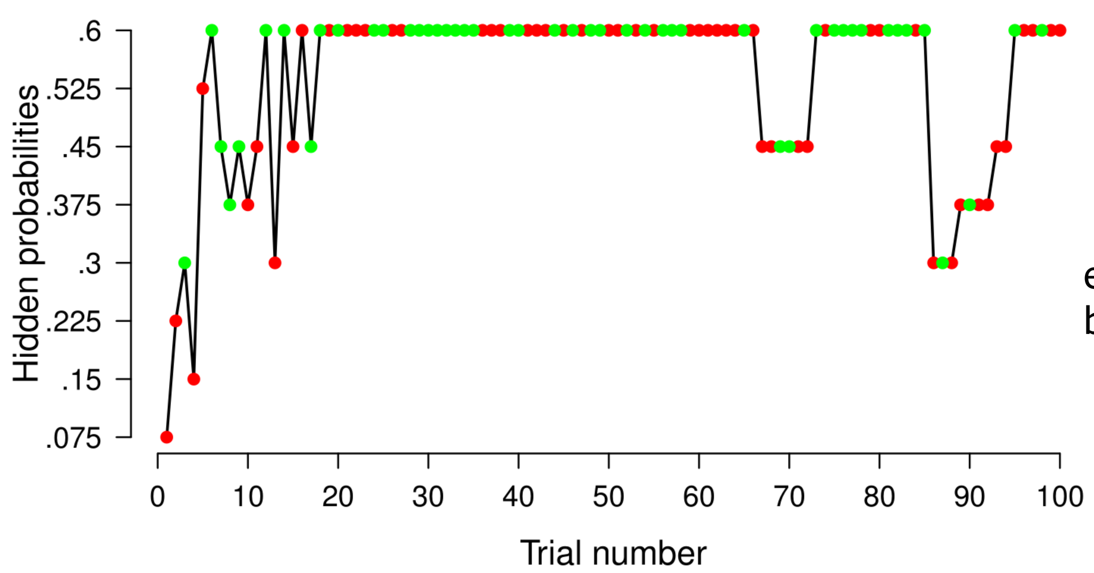
1. OA perform worse in more complex environments
2. OA sub-optimally process cognitively demanding information
3. OA rely more on simpler strategies

Design

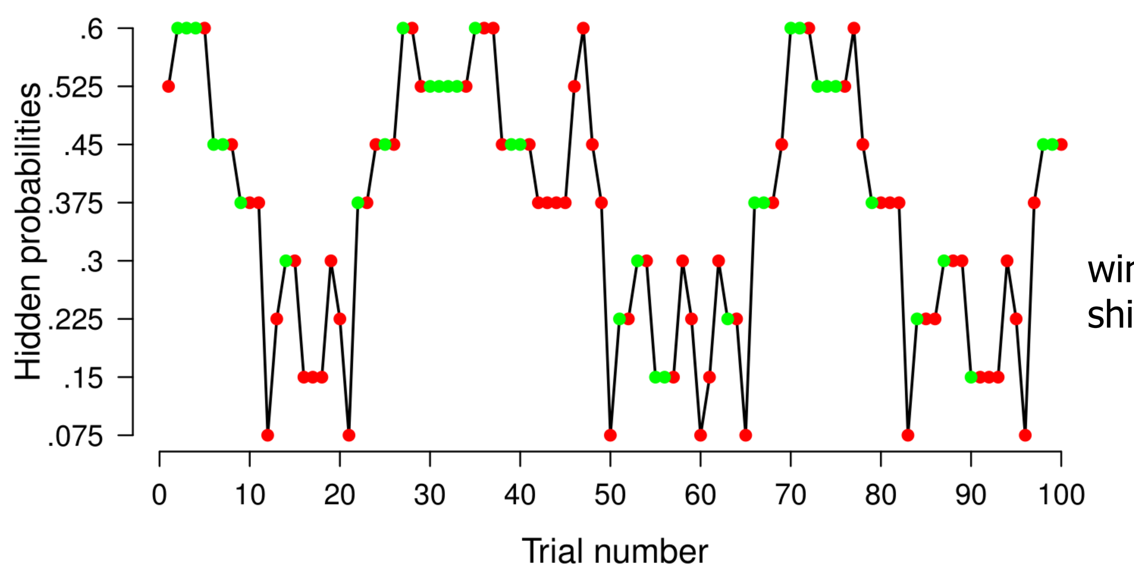
- 4 and 8-armed bandits, binary rewards
- 100 trials, 16 games, ~ 30 min
- Participants are informed that the best bandit has p = .6
- Number of pulls and running averages displayed on each bandit
- 32 YA and 29 OA



Example choice profiles

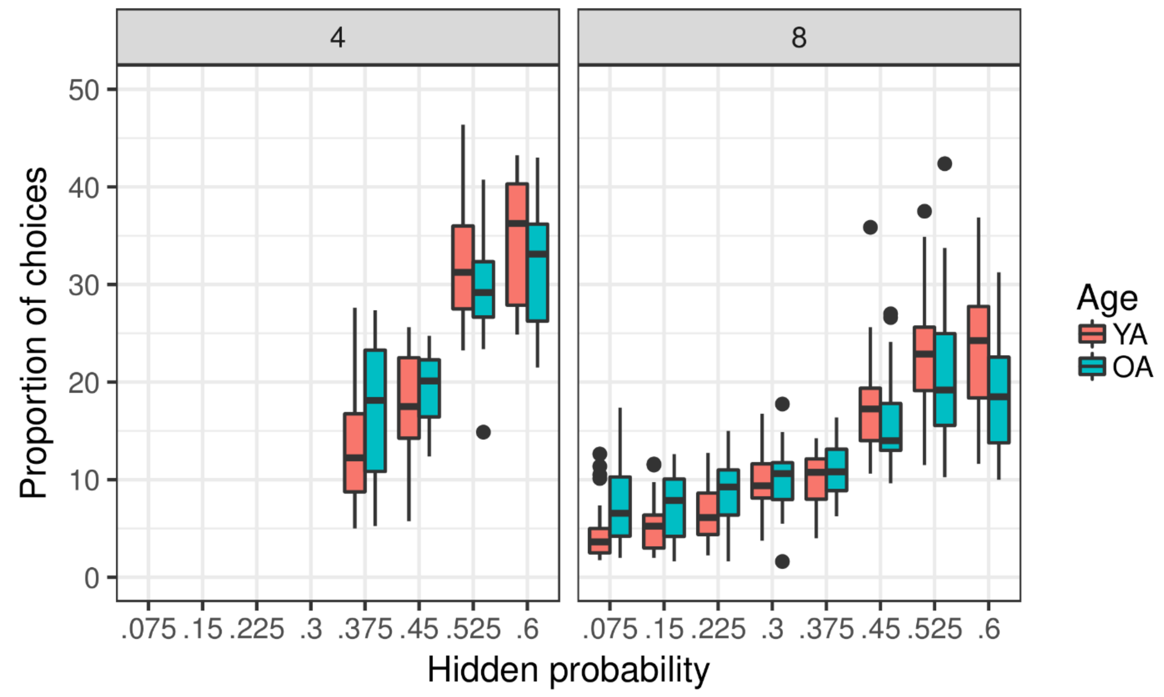


exploration in the beginning



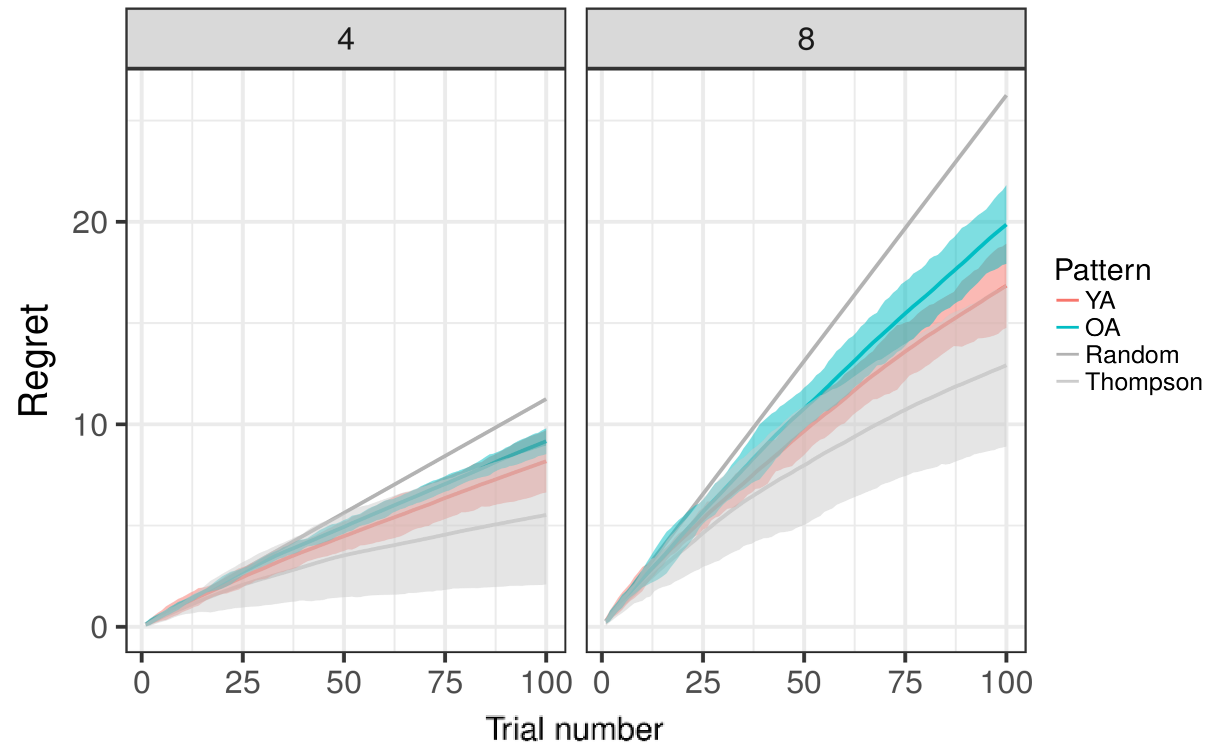
win-stay-lose-shift strategy

Choice proportions



↑OA choose less often from the better options

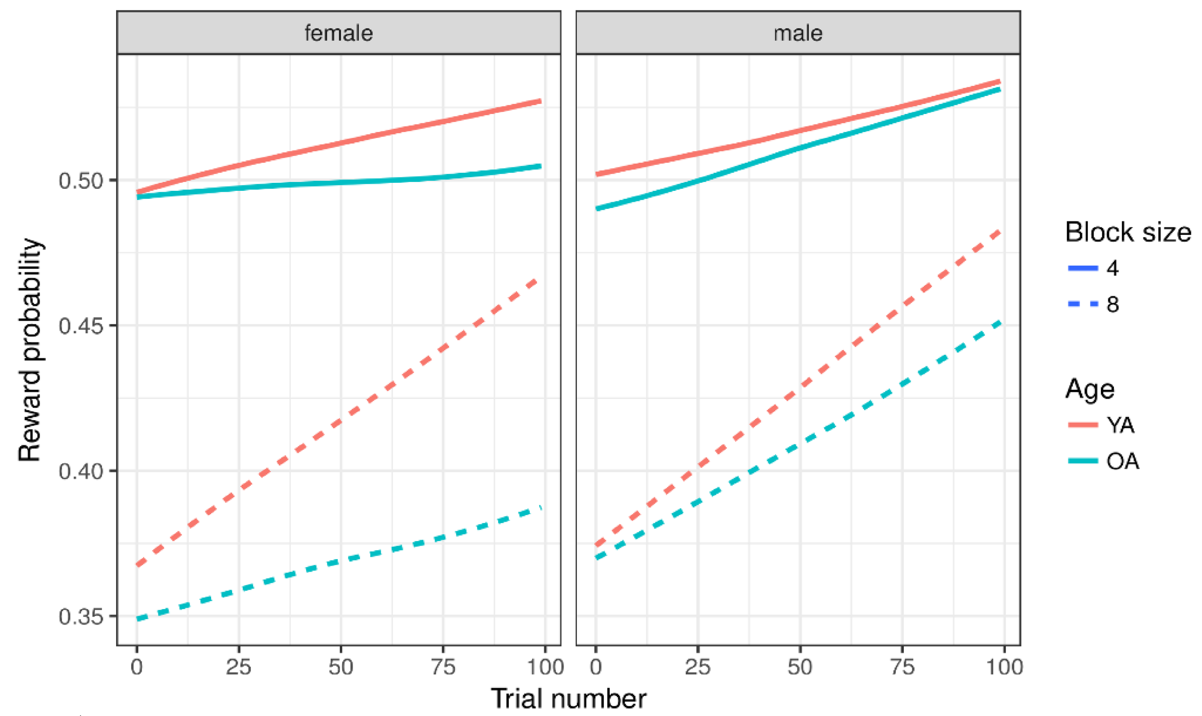
Regret patterns over trials



↑ $R_T = \sum_{i=1}^{100} \left(p_{opt} - p_{B(i)}\right)$, where p_{opt} = .6, and $p_{B(i)}$ is the hidden probability of a chosen bandit

- Random choice regret: 11.25 (4 options), 26.25 (8 options)
- Optimal choice regret: 6.5 (4 options), 11.0 (8 options)

Linear mixed effects regression model

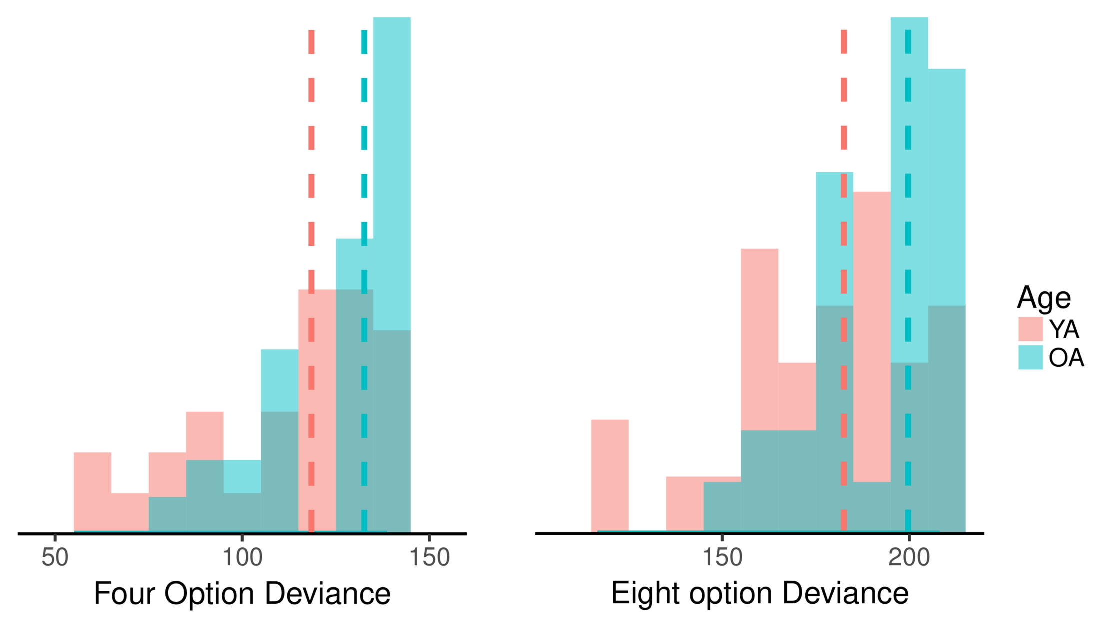


↑
```
hidden p of a chosen bandit ~ (#options + #trial + age)^3 + (1|participant ID) + (1|problem ID) + (1|bandit ID)
```

Three largest effects:

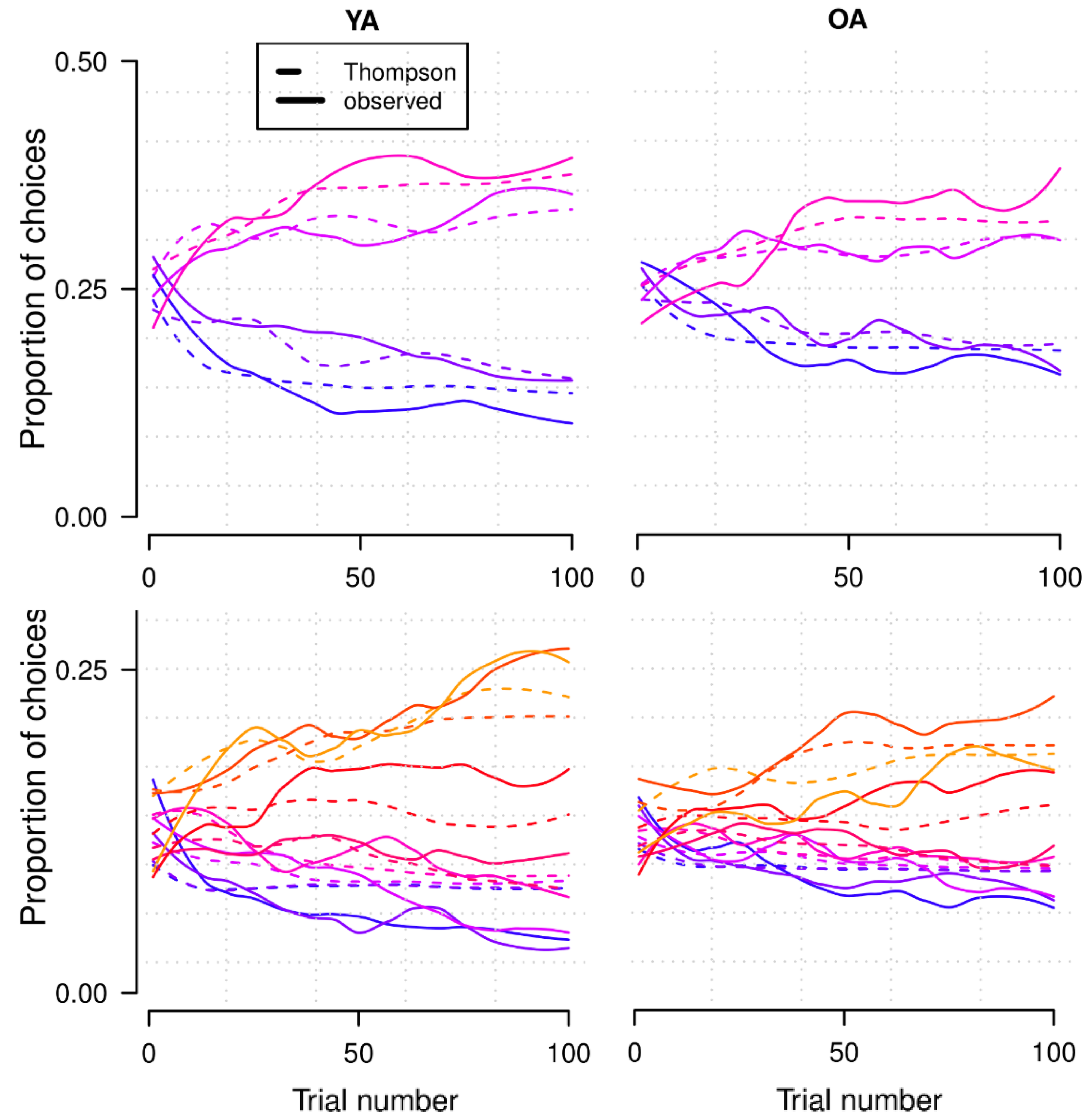
- OA perform worse in 8 options (p < .0001)
- Female OA perform worse in 8 options (p < .0001)
- Both YA and OA improve more in 8 options (p < .0001)

Deviance from fitted Thompson sampling



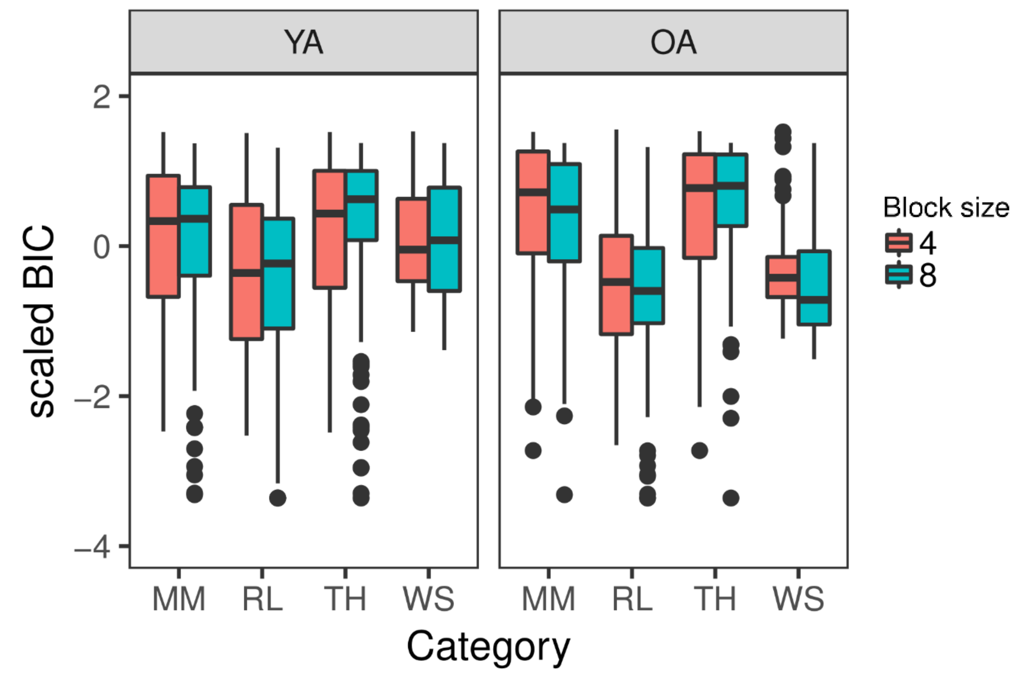
↑Deviance (minimized negative log likelihood) from Thompson sampling using softmax with inverse temperature θ [0.003 - 30].

Choice trajectories



↑Smoothed trajectories of choice proportions and one-step ahead predictions of fitted Thompson sampling. Blue colors represent lower probabilities.

Individual differences



↑BIC statistics for four models across single games, fitted with maximum likelihood and a softmax rule.

- **MM** – always follows the maximum mean
- **RL** – uses a delta rule to update action values
- **TH** – Thompson sampling
- **WS** – Win-stay-lose-shift

Summary

1. OA perform worse in more complex environments
   - OA perform worse in eight options over trials
   - OA explore longer
   - OA have faster increasing regret, leading to a large regret difference for 8 options and medium for 4 options
2. OA sub-optimally process cognitively demanding information
   - OA show larger regret difference to Thompson sampling
   - OA deviated more from fitted Thompson sampling
     - OA inverse temperature θ was also significantly higher
   - OA show less differentiation in one-step-ahead predications
3. OA rely more on simpler strategies
   - OA rely more on win-stay-lose-shift

Conclusions

- Age-differences in explore-exploit tradeoffs depend on task demands
- High performance requires processing of rewards and uncertainty (see e.g. Thompson sampling)
- Why do OA consider such important information less effectively than YA as the choice environment gets more complex?
  - Working memory should be minimal
  - General "slowing", gender effects, more cautious risk taking, etc.
  - Inadequate future reward representations specifically point to under-reliance on goal-directed learning

Acknowledgements

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 github.com/jobschepens/Bandits

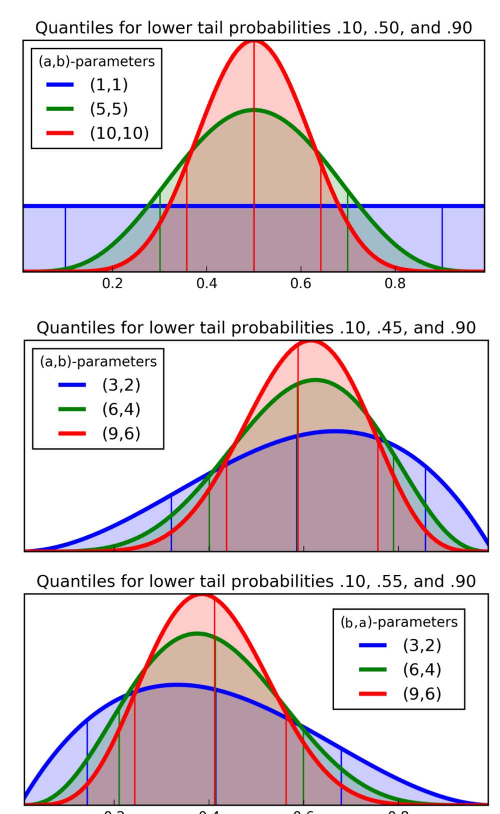
References

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*, 253.
Nassar, M. R., Bruckner, R., Gold, J. I., Li, S. C., Heekeren, H. R., & Eppinger, B. (2016). Age differences in learning emerge from an insufficient representation of uncertainty in older adults. *Nature Communications*, *7*.
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*(4), 1208-1220.

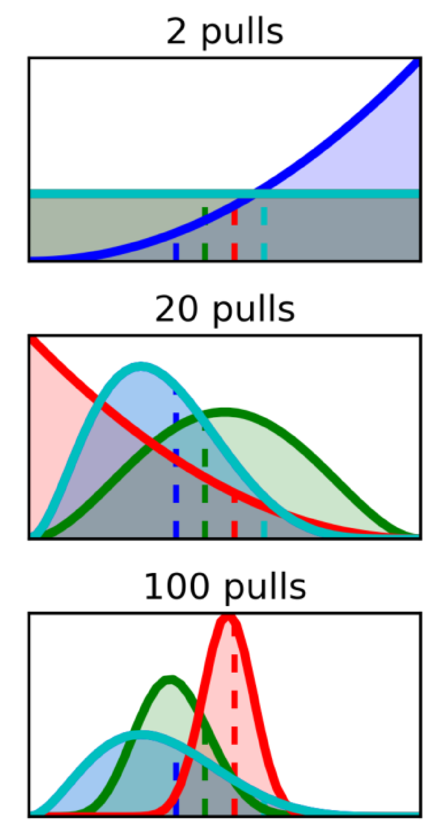
Box 1: Thompson sampling

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

Pseudocode and examples of Thompson sampling, where `n` is the number of options, `x` is a randomly generated probability, and `qbeta` is a function for looking up quantiles from the Beta distribution.



↑How posteriors are updated over time



↑How quantiles compare for different parameter values