Task complexity and experience dictate the use of online, versus offline, planning in humans

Jungsun Yoo and Aaron M. Bornstein

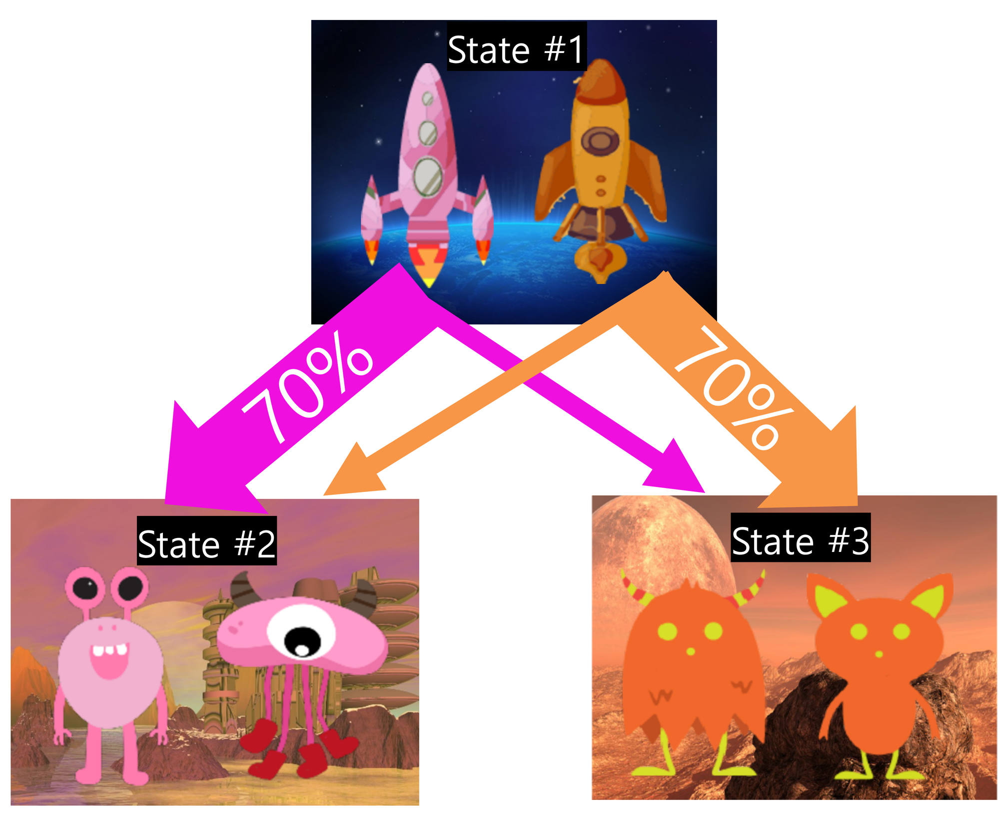
Department of Cognitive Sciences, University of California, Irvine, Irvine, CA 92697 USA

INTRODUCTION

- How much do our decisions depend on **deliberation in the moment**, as opposed to **reasoning performed ahead of time**?
  → **Online** and **offline planning** (Sutton & Barto, 2018)
- Understanding when each kind of planning is used can help predict how decisions may change in response to e.g., time pressures or context effects
- Artificial agents tend to use offline planning, after training on millions of examples (Hamrick et al 2020)
  - But online planning may be more effective when environments are not fully explored, as humans often face
- **Q) Do humans use online, versus offline, planning?**
- Previously, the parameter *w* in two-step tasks (TST) has been used to investigate planning

LIMITATIONS OF TST PARADIGM UPON INVESTIGATING ONLINE PLANNING

1. LACKS TASK COMPLEXITY



- State space complexity (SSC; number of different possible states within a task) is a way to define task complexity (Opheusden et al., 2019)
- In TST, SSC=3 for 100+ trials, and every trial has the same states
- Thus, participants can quickly learn to plan ahead of choices (offline planning), rather than deliberate at each decision time (online planning)

2. *W* CONFLATES ONLINE AND OFFLINE PLANNING

- *w* is a free parameter that captures a subject's tendency to plan:
  $Q = w^{*}Q_{MB} + (1-w)^{*}Q_{MF}$
- However, *w* provides a limited account on planning because it does not differentiate online from offline planning

OUR SOLUTIONS

1. **Increase complexity to increase demands on online planning**
➤ Prediction: Online planning will be more influential when the state spac e is large enough

2. **Measure planning's influence on first-stage response time (OPI: Online Planning Index), as a function of experience**
➤ Prediction: The influence of online planning within planning in general will increase as the agent learns the model

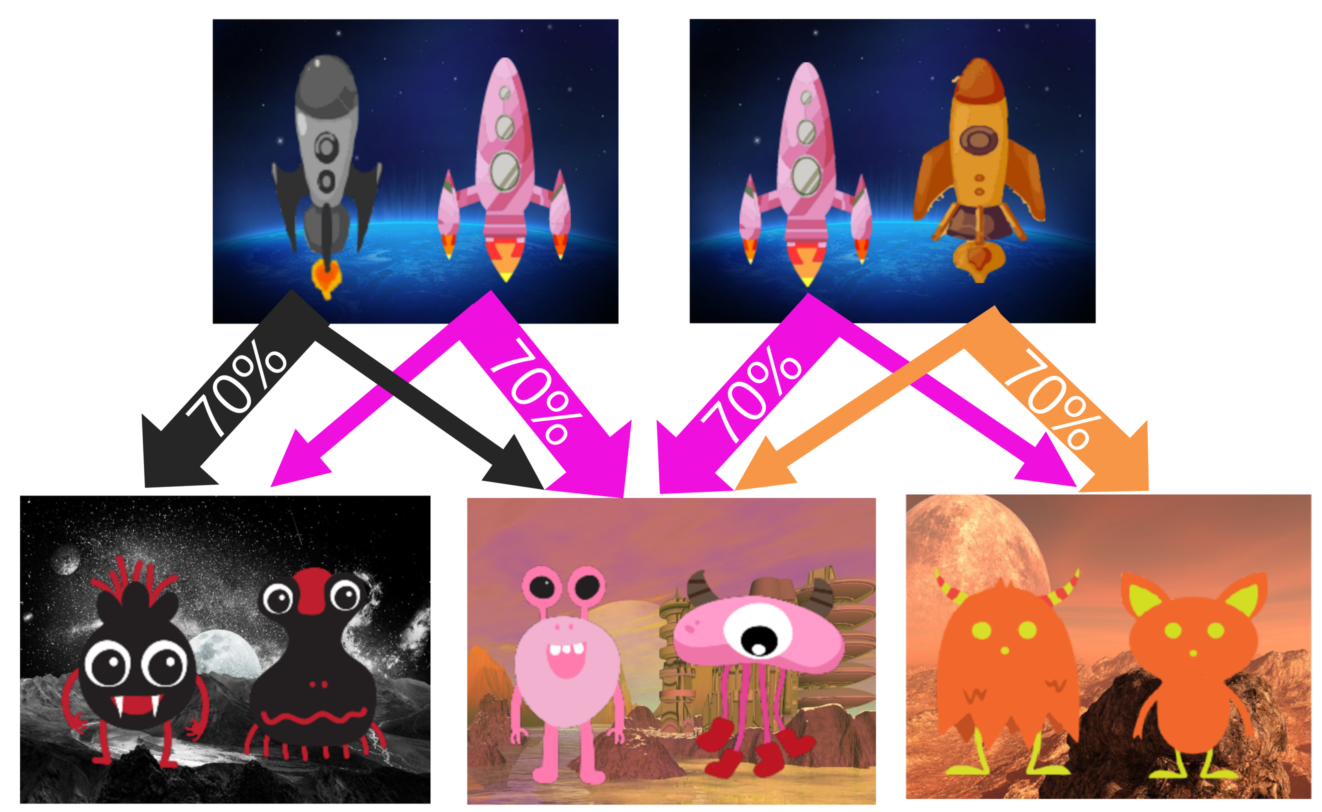
METHODS

1. MULTINOMIAL TST

Trial *n* Trial *n+1*

1[st] stage

2[nd] stage



▪ We varied the possible number of 1[st]-stage options (i.e., spaceships) from **2** (=original TST) to **3, 4,** and **5**
▪ The transition structure varied dynamically according to the combination of the two spaceships presented
▪ Thus, *n* spaceships yield a total of ${}_{n}C_{2}$ (**1[st] stage) + *n* (2[nd] stage)** SSC: **3, 6, 10, and 15 SSC** for 2-, 3-, 4-, and 5-spaceship TST
▪ We conducted the four TST variants on N=110 participants each

2. ONLINE PLANNING INDEX

▪ OPI: an operational measure of decision-time deliberation
- The degree to which model-based values (Q_{MB}) explain response times in first stage choices (RT_1)

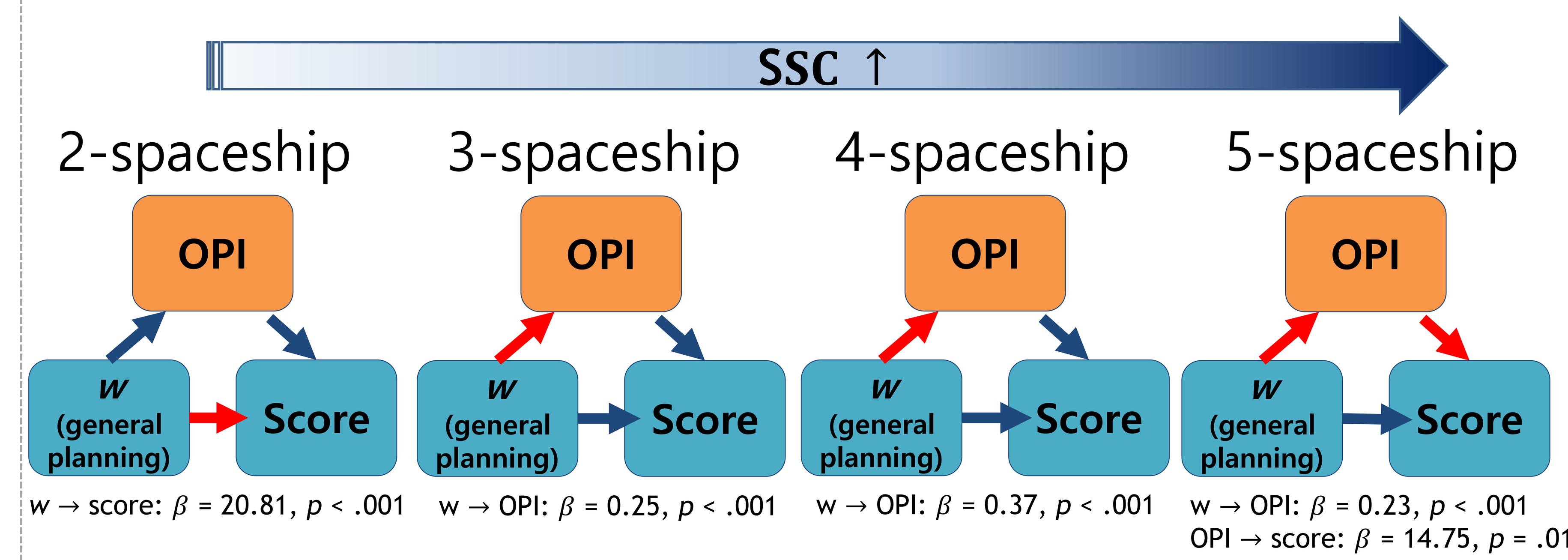
▪ Assumption: options with similar values lead to longer RTs
▪ Formula:

$$RT_1 = \beta_{MB} \times (Q^1_{MB} - Q^2_{MB}) + \beta_{MF\,\times}(Q^1_{MF} - Q^2_{MF})$$

$$OPI = \frac{1-\beta_{MB}}{1-\beta_{MF}}$$

RESULTS

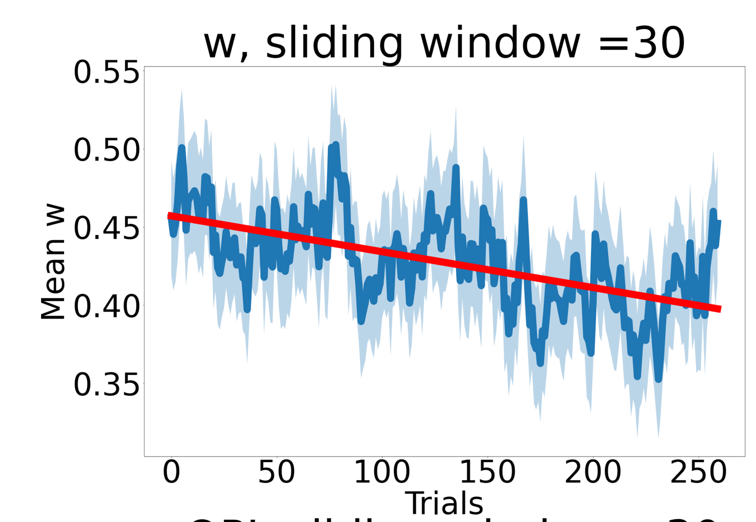
1. Participants were more likely to use online planning as a function of **SSC**

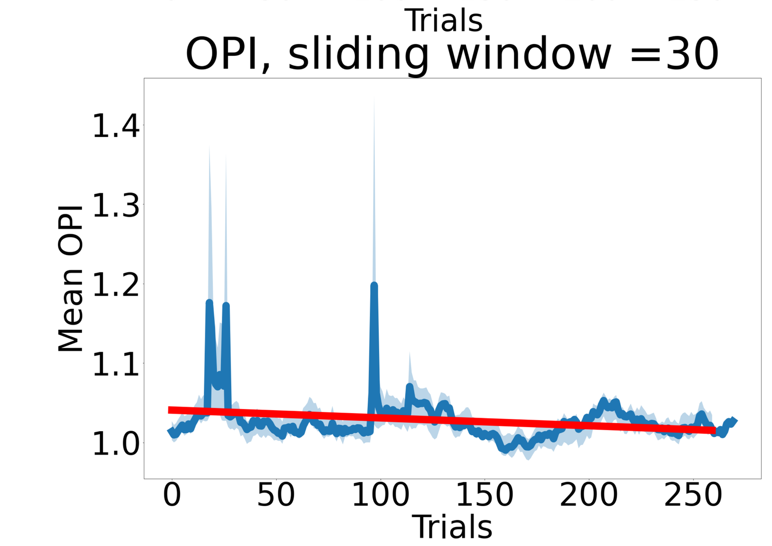


▪ The planning tendency in general (*w*) was directly related to performance in the canonical TST
▪ In more complex variants this relationship grew progressively more mediated by the degree of online planning, to the extent that
▪ **A mediation effect was observed in the 5-spaceship TST**

2. The influence of online planning increases within general planning, as a function of **within-task experience**

Within – task experience ↑





▪ Consistent with uncertainty-weighted arbitration theory, model-based influence on choice decreases with experience

▪ However, consistent with previous observations that w measures model use, but not model learning (Konavalov & Krajbich 2020), OPI remains steady across experience

▪ These results suggest that 1) OPI could be a more stable measure of model use (vs. model learning), and 2) since OPI remains constant while general planning tendency decreases, its influence vs. offline planning increases with experience

CONCLUSIONS

▪ These results suggest that humans learn to use online planning in complex environments.
▪ Our findings have implications for the generalizability of decision-making models to naturalistic settings, and for the correspondence between decision-making in humans and artificial agents.

CONTACT: Jungsun Yoo (jungsun.yoo@uci.edu)