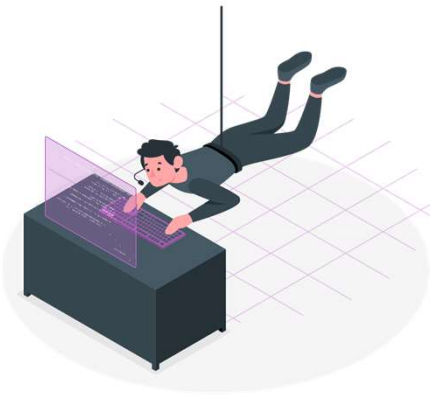
Humans monitor learning progress in curiosity-driven exploration

Ten, A., Kaushik, P., Oudeyer, P. Y., & Gottlieb, J. (2021). Humans monitor learning progress in curiosity-driven exploration. *Nature communications*, 12(1), 1-10.

Presented by Helena Luo 08/11/2022

The strategic learner

- How learners allocate study time to maximize learning across a set of the activities



How to self organise exploration?

- Curiosity – desire to know, value for information independent of material gains
- Operationalised as intrinsically motivated information demand show preference for information (encoded in neural systems of reward and motivation)
- But most have short time scales - how do people self-organise exploration to learn on longer time scales?

Where to allocate time?

- Optimal allocation for ‘strategic student’ is very sensitive to the shape of the expected learning trajectory
  - Concave LC – optimal to spend longer on lower levels of competence (foundation)
  - Logistic LC - optimal allocations vary with time availability
- But this is not trajectory is not available to learners in practice

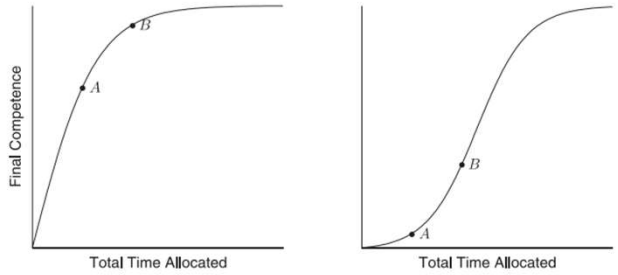
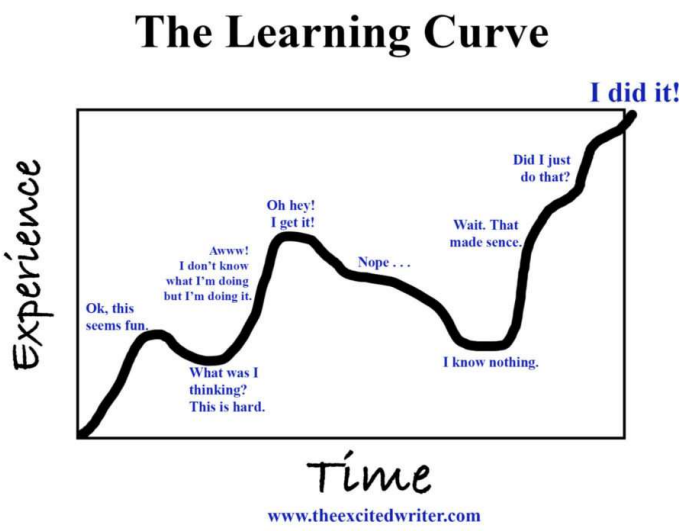


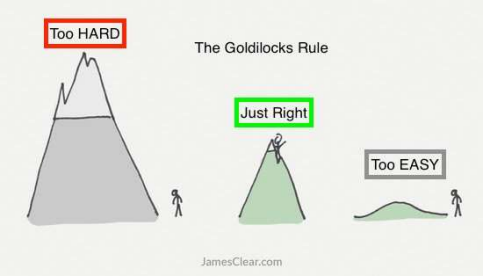
Fig. 1. Examples of learning curves.



Prioritise study on perceived difficulty

- Preference to prioritise high difficulty vs intermediate difficulty
- High difficulty – computational architecture that assigns intrinsic utility for prediction errors or uncertainty
- Intermediate difficulty – control architectures using learning progress (LP)
  - LP is measure of percentage correct (PC) over certain timeframes
  - LP is change in performance while PC is absolute performance (LP is temporal derivative of performance)

Control using learning progress



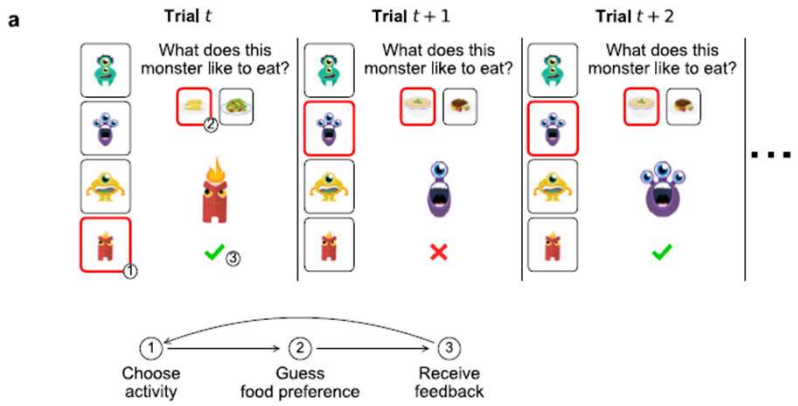
- PC-based algorithms steer agents towards hardest tasks
- LP-based algorithms important in naturalistic environments as it steers agents away from highly familiar tasks and also unlearnable tasks (e.g. intrinsically random or cannot be mastered with current knowledge or skills)
  - Applied to Automatic Curriculum Learning (ACL, family of mechanism in Deep Reinforcement Learning)
  - Used to personalize sequences of learning activities in educational technologies

- "Despite the potential importance of LP-based control strategies, there is no empirical evidence of whether, and how, people use such strategies"
  - Previous studies measured difficulty based on familiarity with the topic, "no study has tested whether participants can dynamically monitor their performance on an arbitrary activity and use dynamic estimates of PC or its temporal derivative (LP) as predicted by computational algorithms"

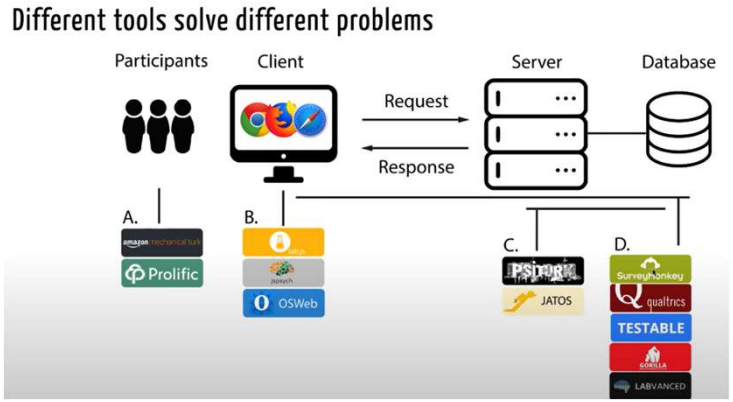
Monster feeding task

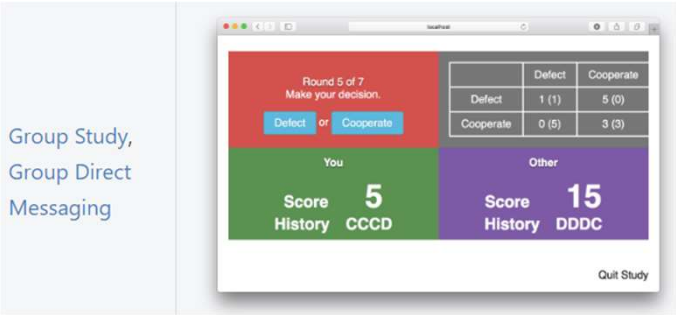
- How people self-organize their exploration over a set of activities of variable difficulty
- Whether they spontaneously adopt learning maximization objectives when they do not receive explicit instructions

- Familiarisation (4 x 15 forced-choice), subjective rating, 250 trials of free-choice, subjective rating
- Vary in complexity (within subject)
  - Dimension and relevance
  - A1 (1d1), A2 (2d1), A3 (2d2), A4 (2d0) i.e. random
- Vary in learning objectives (between subject)
  - EG – external goal (N = 196) – were asked to maximize learning across all the activities and were told that they will be tested at the end of the session (reward not tied to performance)
  - IG – internal goal (N = 186) – choose what they prefer
- Recruited via Amazon Mechanical Turk, time taken about 45 min to 60 mins, upon completion, compensated $1 regardless of performance
  - "was consistent with prevailing rates on Amazon MTurk and with our goals of minimizing the role of monetary incentives and avoiding biasing participants toward activities with consistently high performance"
  - "rather than rewarding participants for individual correct answers, our external instruction specified the end-goal but not the local strategy for achieving the goal; this allowed people to choose their activities and commit errors in the short term, in the interest of maximizing learning in the long term. This greater autonomy, we believe, contributed to the synergism we observed, whereby externally imposed goals enhanced the eventual learning outcomes, rather than hindering them"



Side note about experimental methods 1





- Conducted on JATOS (Just Another Tool for Online Studies)
  - Free and open source
  - Run studies on your own server (e.g. at your university) – keep complete control over who can access your result data and can comply with your ethics
  - Run group studies where multiple participants interact with each other in real-time

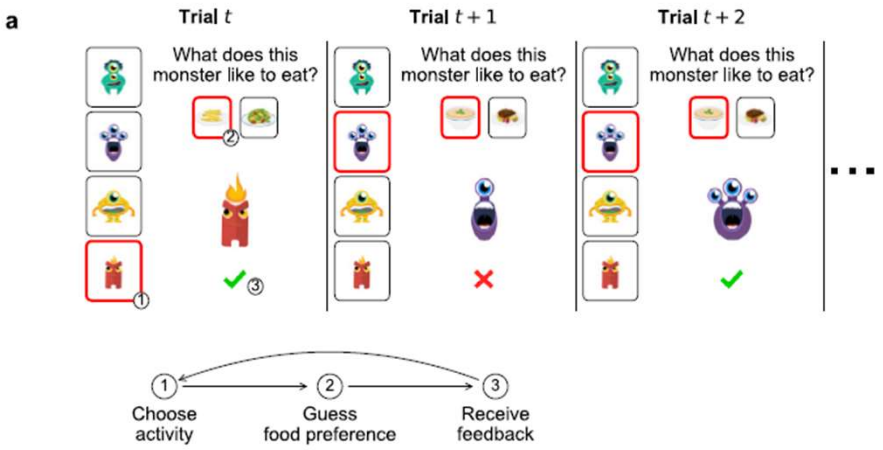
Side note about experimental methods 2

- Random shuffling of stimulus
- Dichotomous choice in original study but will use confidence interval in new study
  - Must follow instructions to pass and start the task – encouraged to use slider scale
- Free choice task (and really engaging experimental design)

Monster feeding task

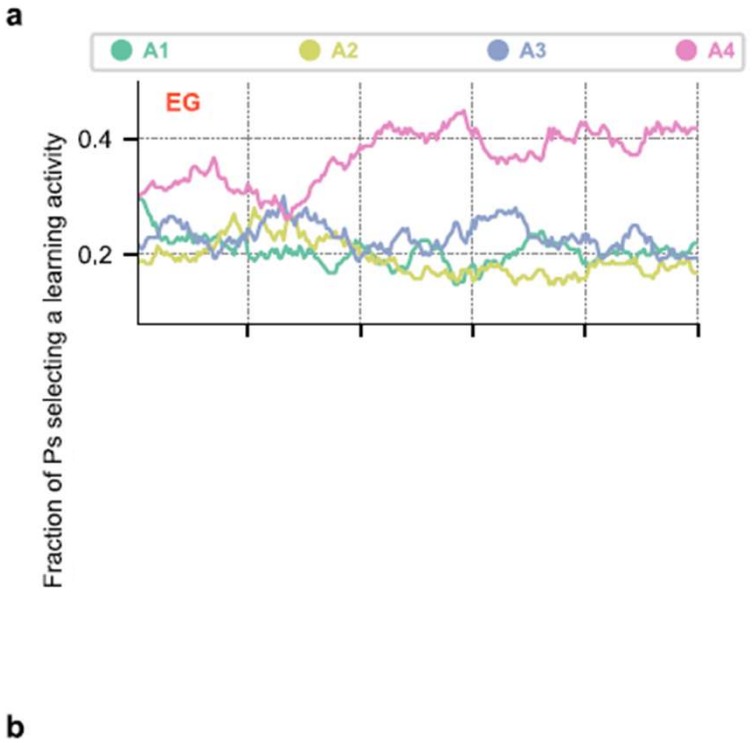
- How people self-organize their exploration over a set of activities of variable difficulty
- Whether they spontaneously adopt learning maximization objectives when they do not receive explicit instructions

- Vary in complexity (within subject)
  - Dimension and relevance
  - A1 (1d1), A2 (2d1), A3 (2d2), A4 (2d0) i.e. random
- Vary in learning objectives (between subject)
  - EG – external goal (N = 196) – were asked to maximize learning across all the activities and were told that they will be tested at the end of the session (reward not tied to performance)
  - IG – internal goal (N = 186) – choose what they prefer



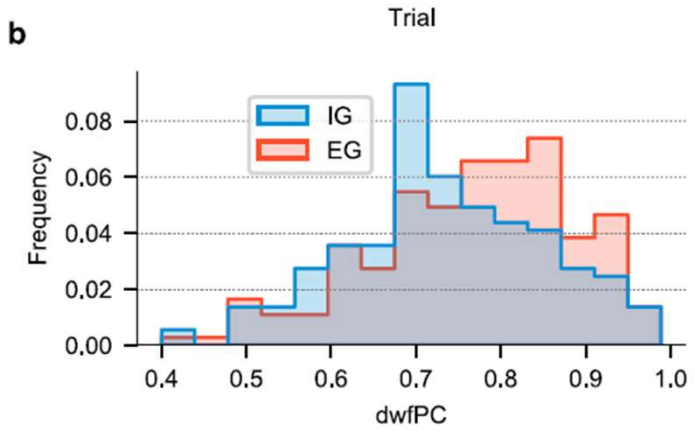
Activity choice varies slightly by group

- EG and IG had same performance in familiarisation stage but different choice patterns in free-choice
- EG group focused strongly on the most difficult activity (the unlearnable activity that had the lowest PC, 37%)
- IG group showed a more uniform preference with only a slight bias toward the easiest activity (A1: 33%)
- This difference was significant using ANOVA of time allocation



Average learning varies slightly by group

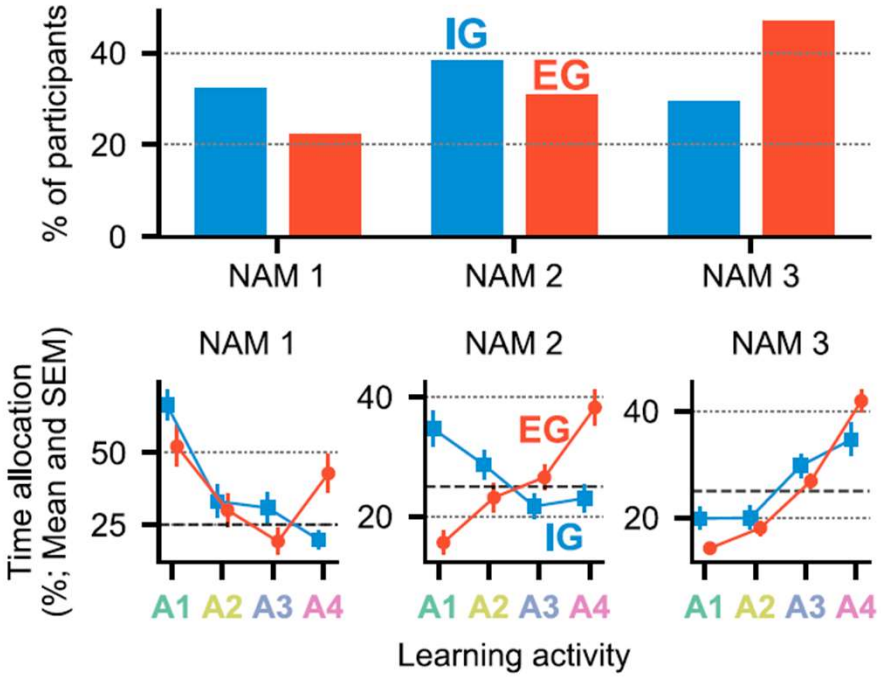
$$\mathrm{dwfPC}_i = \tfrac{1}{6}\mathrm{fPC}_{i,A1} + \tfrac{1}{3}\mathrm{fPC}_{i,A2} + \tfrac{1}{2}\mathrm{fPC}_{i,A3}$$



- Difficulty weighted final PC (dwfPC)
- The average PC in the last 15 trials spent on each activity scaled by its difficulty rank
- Slightly higher for EG (M = 0.756, SD = 0.127) relative to the IG group (M= 0.721, SD = 0.126)

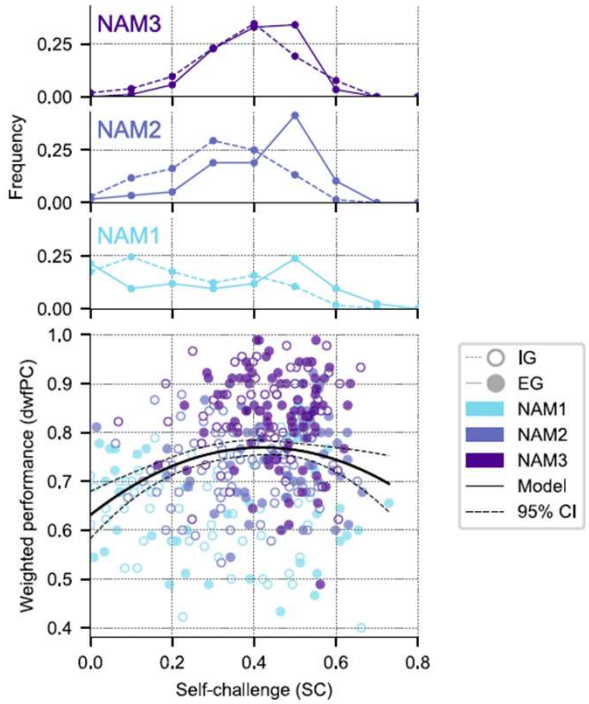
Substantial individual variability in subset of IG: number of activities mastered (NAM)

- Categorised each participant based on the number of activities they mastered
  - Criterion of 13/15 correct trials (86.7% correct), binomial distribution of p = 0.0037 of arising by chance
- Number of participants:
  - 65% of IG participants mastered more than one activity (NAM2 and NAM3) and 30% mastered all 3 activities
  - Comparable to EG group, where 74% mastered at least 2 activities, and 46% mastered all three.
- Time spent:
  - IG - NAM1 and NAM2 showed choices consistent with the group average (favoring the easiest activity), NAM3 participants showed a distinct preference for A3 and A4 activities that more closely resembled the EG group
  - NAM1 and NAM2 groups differed in activity selection across the instruction conditions (EG vs IG)
  - NAM 3 allocated their time similarly - a sizeable fraction of the IG group behaved in the same way as people who were instructed to learn and prepare for a test



Self-challenge index

$$SC_{t,i} = 1 - \frac{PC_{t,i} - \min_{\forall k \in K} PC_{:t,k}}{\max_{\forall k \in K} PC_{:t,k} - \min_{\forall k \in K} PC_{:t,k}}$$

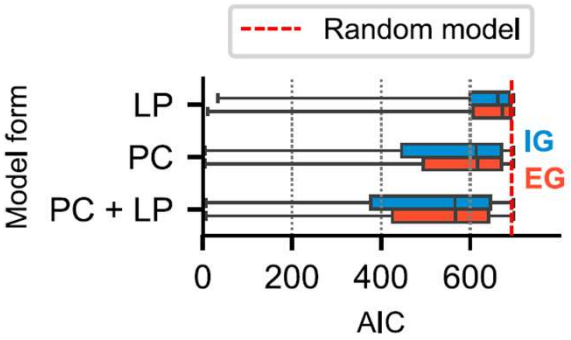


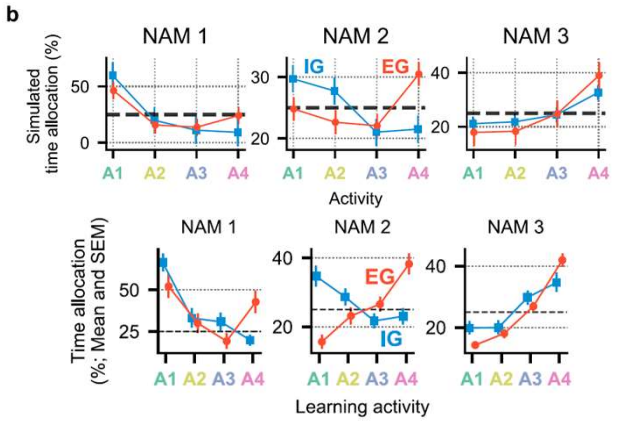
- Recent PC of the activity selected on each trial, normalized to the entire range of PC levels the participant experienced so far
- 0 = tended to choose the easiest of the activities they experienced, 1 = choose the most difficult activities
- Best learning outcomes were associated with intermediate SC – dwfPC has a strong inverted-U relationship with SC
- Participants with different instructions and learning achievement fell on different portions of the inverted-U curve:
  - Did not master all 3 activities (NAM1 and NAM2) fell on the rising and falling arms of the inverted-U curve:
    - EG participants who failed to master all 3 tasks did so because they over-challenged themselves and those in the IG group did so because they under-challenged themselves
  - Mastered all 3 activities were at the top of the inverted-U curve and had equivalent (intermediate) SC in the IG and EG groups
    - Consistent with the activity preferences – a subset of participants spontaneously adopted intermediate self-challenge strategies and maximized learning regardless of external instructions

Dynamic sensitivity to LP

- Monitor change in performance over time in combination with PC
- Utility of an activity is a linear combination of PC and LP
  - Fit data to probabilistic (softmax) choice over 4 discrete classes, using MLE with 3 free parameters:
    - softmax temperature (capturing choice stochasticity)
    - Weights wPC, wLP as sensitivity respectively, PC and LP
- Model with PC and LP provided superior fit
  - Lack of difference between IG and EG - do not need to be explicitly instructed to maximize learning to demonstrate sensitivity to LP
- Validated using simulations of time-allocation using the coefficients fitted by the bivariate models similar to actual behaviour

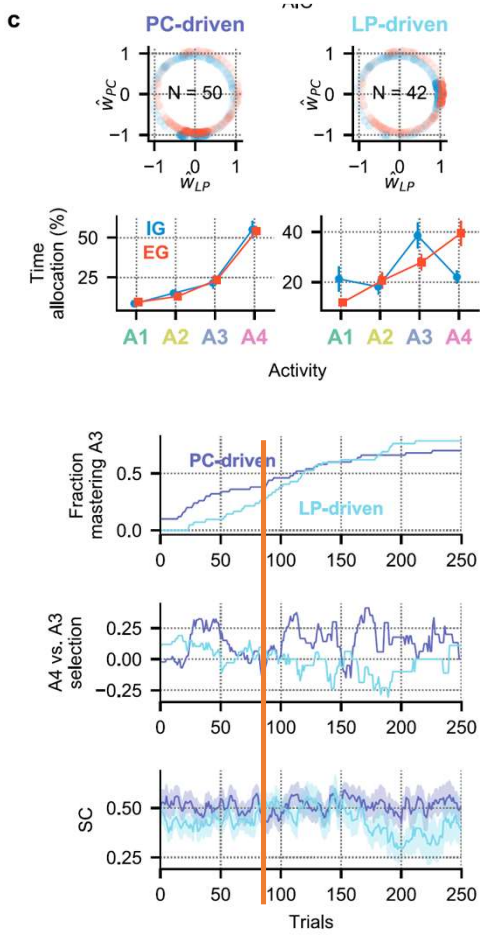
$$U_{i,t} = w_{\mathrm{PC}} \times \mathrm{PC}_{i,t} + w_{\mathrm{LP}} \times \mathrm{LP}_{i,t}$$





LP sensitivity to steer away from too hard

- Focused on two subsets of participants whose choices were driven predominantly PC or LP
- Both groups preferred more difficult activities
- Preference for A4 was lower in LP-driven relative to PC-driven participants
- Lower preference for A4 enhanced learning outcomes in the LP-driven relative to the PC-driven group (increase in fraction mastering A3)
  - The probability of mastering at least 2 activities was 90.48% in the LP-driven group versus 70.59% the PC-driven group, and the probability of mastering all 3 tasks was, respectively, 64.29% versus 34.98%.
  - Thus, consistent with theoretical predictions, LP-driven choices increase the efficiency of active learning by steering participants away from unlearnable activities.



Key findings

How people self-organize their exploration over a set of activities of variable difficulty?

- Preference for intermediate complexity extends to choices of learning activities (due to underlying LP mechanism)
- Learner's competence (prediction errors or error rates) and changes in competence over time (learning progress) jointly shape activity choice and exploration
  - PC – preference for high errors – explore harder unfamiliar activities
  - LP – preference for temporal derivative of PC – avoids unlearnable activities

Whether people spontaneously adopt learning maximization objectives when they do not receive explicit instructions?

- Intrinsic and extrinsic motivation is complex
  - EG more likely to self challenge (greater tolerance for errors, better overall learning, but made some labour in vain in unlearnable activity)
  - How best to balance EG and IG for efficient learning in certain context?
- Longer time scale – momentary curiosity vs sustained interest
  - Four stage model of interest development: situational interests is initially triggered and sustained (or dampened) by the environment but with time gives way to well developed interest in which people spontaneously generate new questions and initiate investigations
  - Many people in IG group mastered two or more tasks and reported subjective interest proportional to their time allocation suggests that the activities may have triggered their situational interest regardless of explicit instructions
  - Higher achievements in the EG group suggests that external instructions help support that fledgling interest