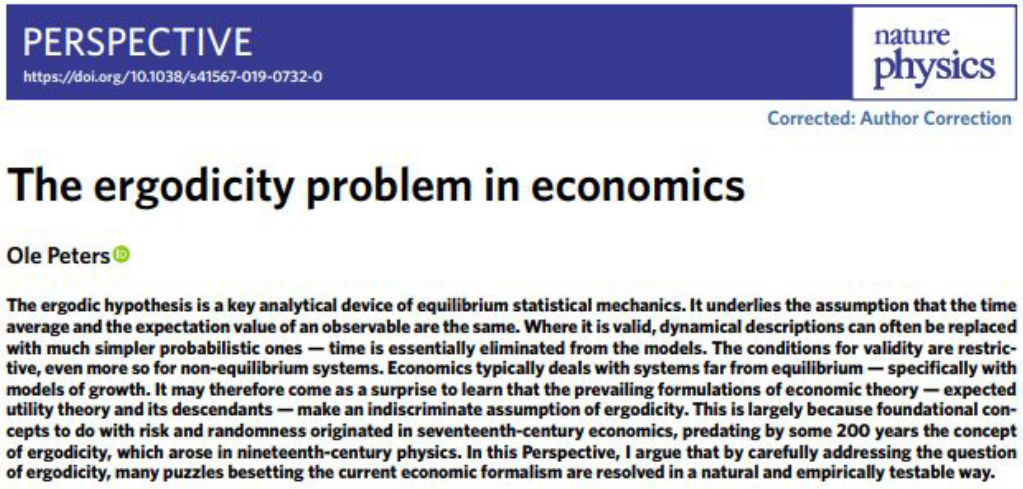
Ergodicity breaking in Reinforcement Learning

When expected values are not the value you expect

Bert Verbruggen, Arne Vanhoyweghen, Vincent Ginis

Why did we want to study this question?



PERSPECTIVE
https://doi.org/10.1038/s41567-019-0732-0

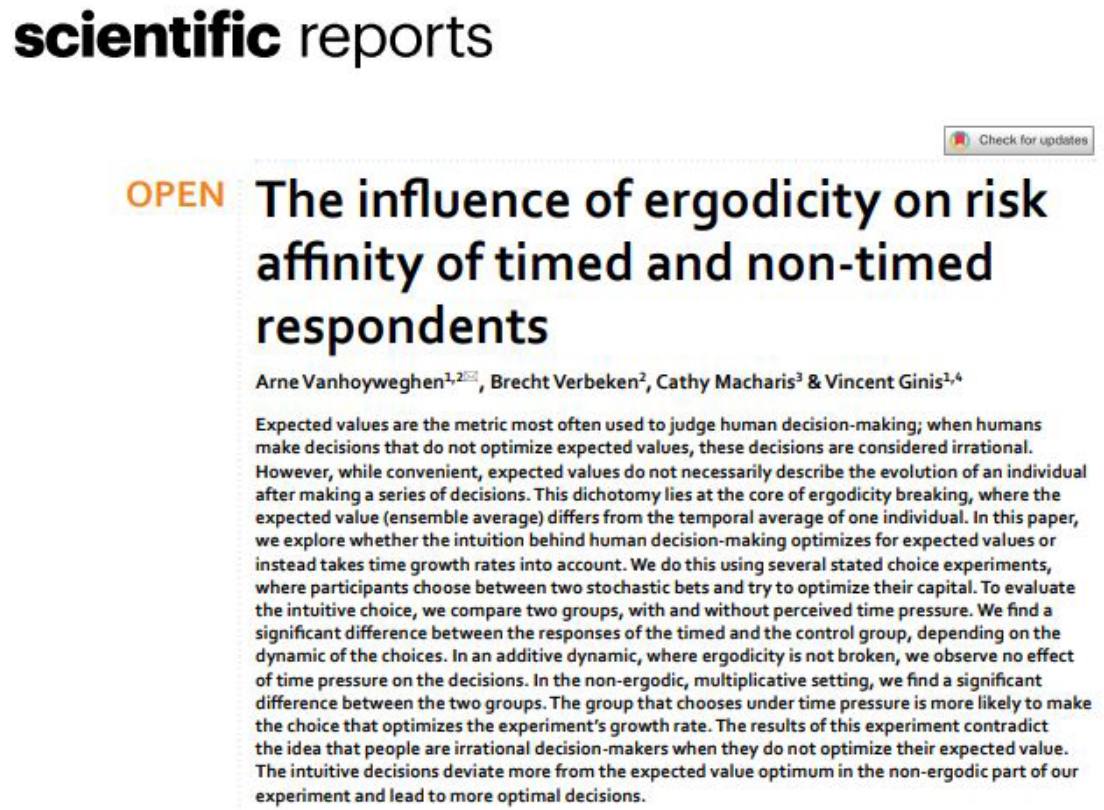
nature physics

Corrected: Author Correction

The ergodicity problem in economics

Ole Peters

The ergodic hypothesis is a key analytical device of equilibrium statistical mechanics. It underlies the assumption that the time average and the expectation value of an observable are the same. Where it is valid, dynamical descriptions can often be replaced with much simpler probabilistic ones — time is essentially eliminated from the models. The conditions for validity are restrictive, even more so for non-equilibrium systems. Economics typically deals with systems far from equilibrium — specifically with models of growth. It may therefore come as a surprise to learn that the prevailing formulations of economic theory — expected utility theory and its descendants — make an indiscriminate assumption of ergodicity. This is largely because foundational concepts to do with risk and randomness originated in seventeenth-century economics, predating by some 200 years the concept of ergodicity, which arose in nineteenth-century physics. In this Perspective, I argue that by carefully addressing the question of ergodicity, many puzzles besetting the current economic formalism are resolved in a natural and empirically testable way.



scientific reports

Check for updates

OPEN

The influence of ergodicity on risk affinity of timed and non-timed respondents

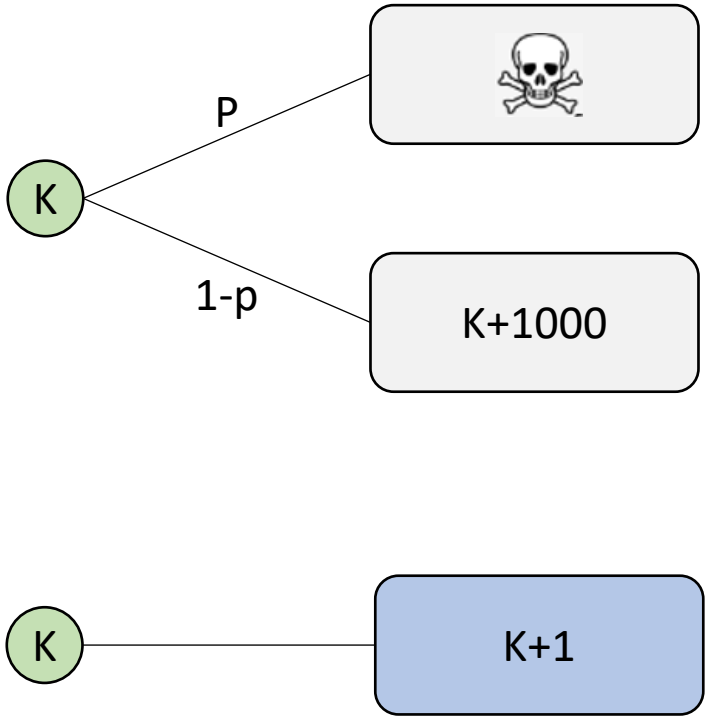
Arne Vanhoyweghen[1,2], Brecht Verbeken[2], Cathy Macharis[3] & Vincent Ginis[1,4]

Expected values are the metric most often used to judge human decision-making; when humans make decisions that do not optimize expected values, these decisions are considered irrational. However, while convenient, expected values do not necessarily describe the evolution of an individual after making a series of decisions. This dichotomy lies at the core of ergodicity breaking, where the expected value (ensemble average) differs from the temporal average of one individual. In this paper, we explore whether the intuition behind human decision-making optimizes for expected values or instead takes time growth rates into account. We do this using several stated choice experiments, where participants choose between two stochastic bets and try to optimize their capital. To evaluate the intuitive choice, we compare two groups, with and without perceived time pressure. We find a significant difference between the responses of the timed and the control group, depending on the dynamic of the choices. In an additive dynamic, where ergodicity is not broken, we observe no effect of time pressure on the decisions. In the non-ergodic, multiplicative setting, we find a significant difference between the two groups. The group that chooses under time pressure is more likely to make the choice that optimizes the experiment's growth rate. The results of this experiment contradict the idea that people are irrational decision-makers when they do not optimize their expected value. The intuitive decisions deviate more from the expected value optimum in the non-ergodic part of our experiment and lead to more optimal decisions.

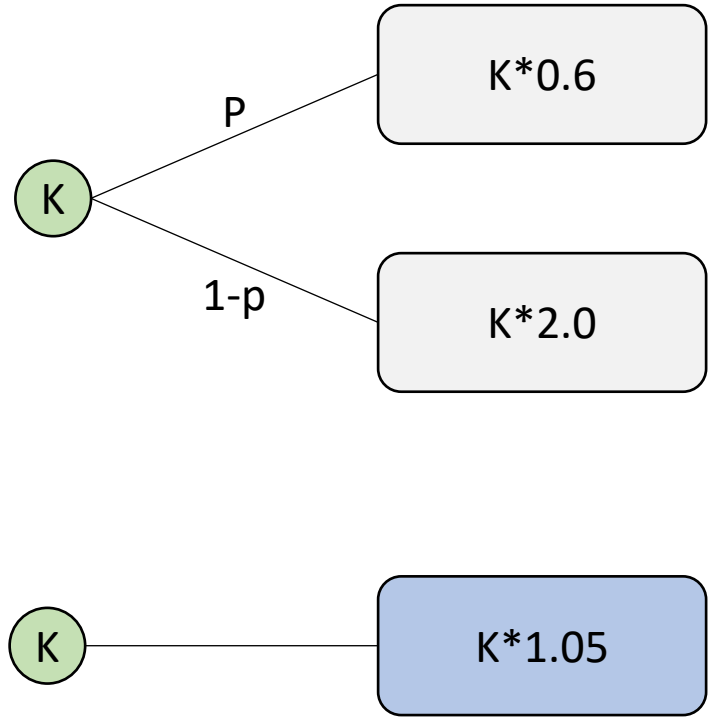
Vanhoyweghen, A., Verbeken, B., Macharis, C., & Ginis, V. (2022). The influence of ergodicity on risk affinity of timed and non-timed respondents. Scientific reports, 12(1), 1-9.
Peters, O. (2019). The ergodicity problem in economics. Nature Physics, 15(12), 1216-1221.

Do you want to play a game?

Additive game



Multiplicative game

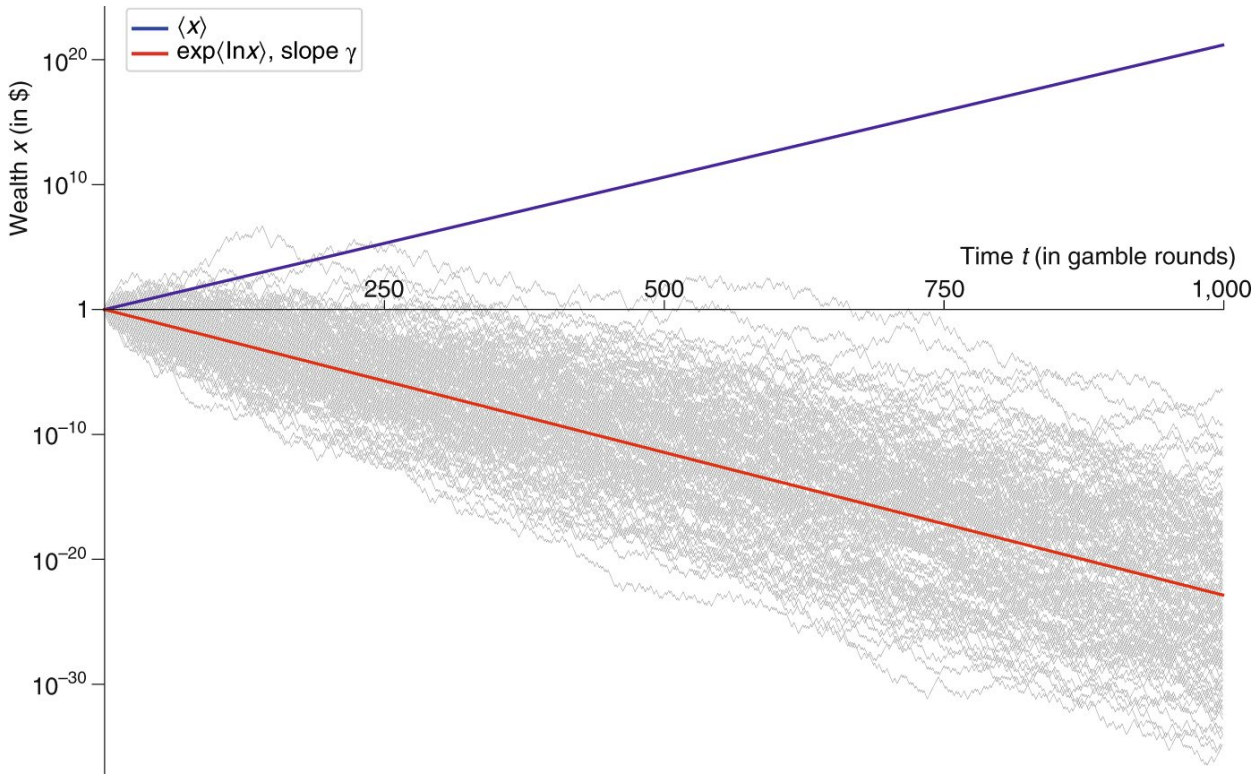


Ergodicity breaking: Main message

- Birkhoff ergodicity theorem:

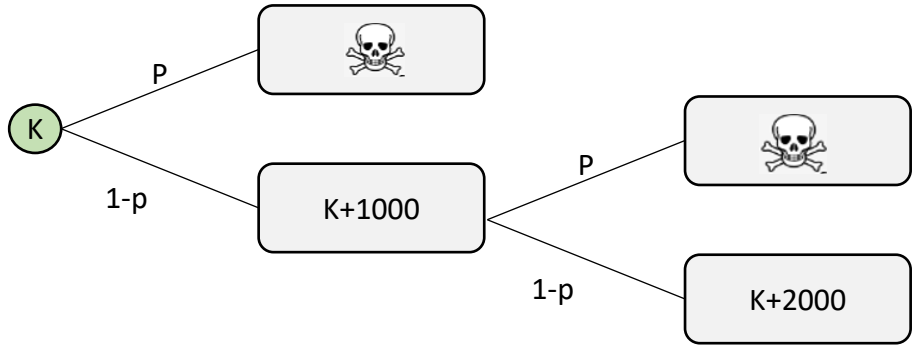
$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f\left(\omega\left(t\right)\right) dt = \int_\Omega f\left(\omega\right) P\left(\omega\right) dt$$

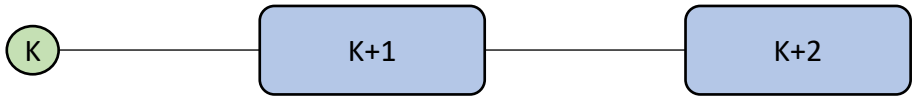
- Group or ensemble average is not the same as time average
- Result for training different agents once on the problem differs from one agent trained repeatedly



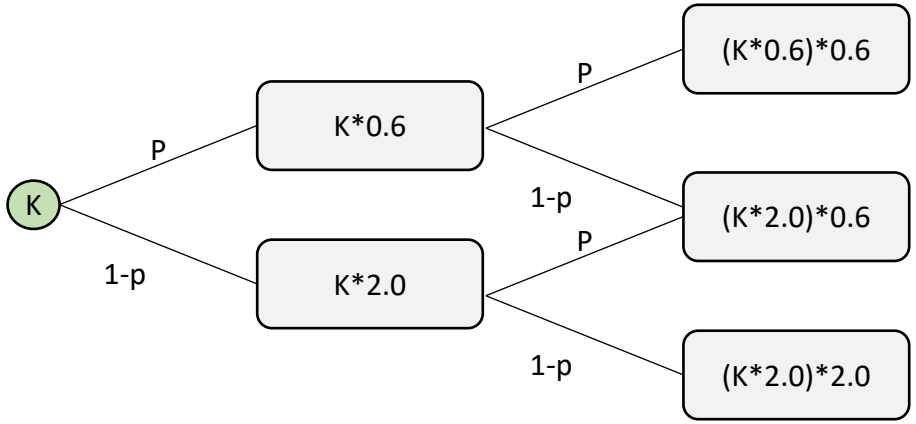
Do you want to play a game? Again?

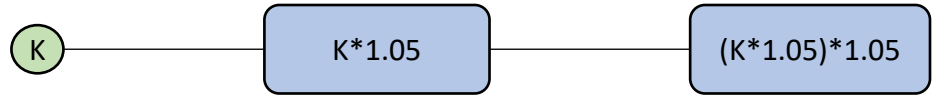
Additive game





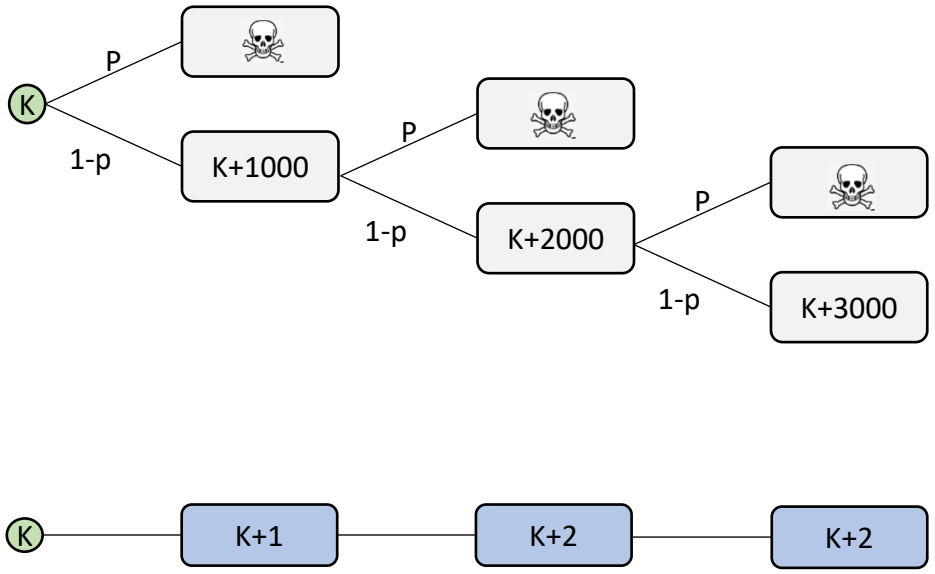
Multiplicative game



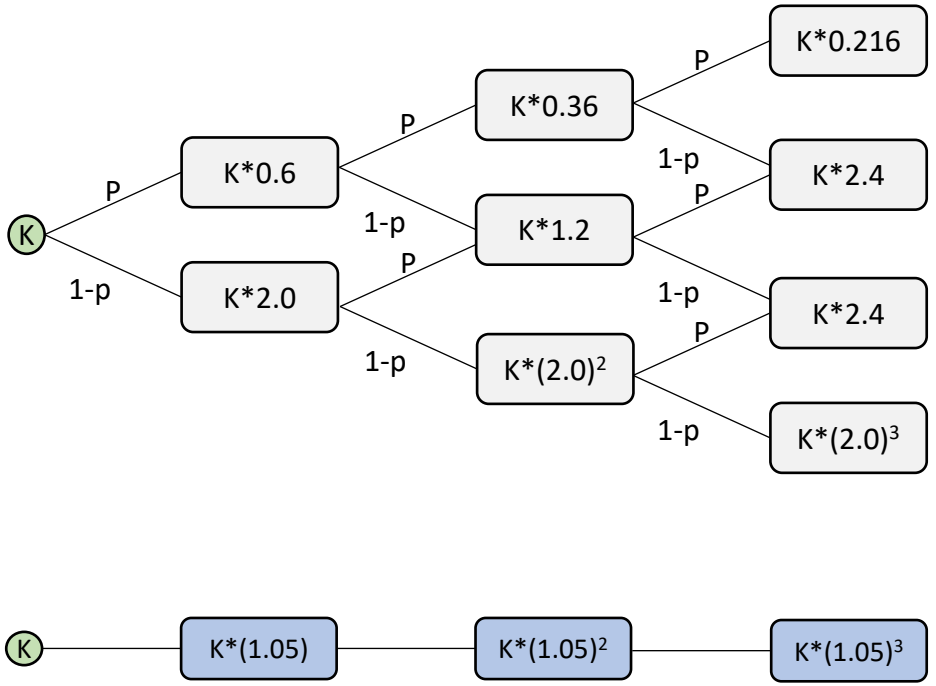


Do you want to play a game?
Again, and again, …

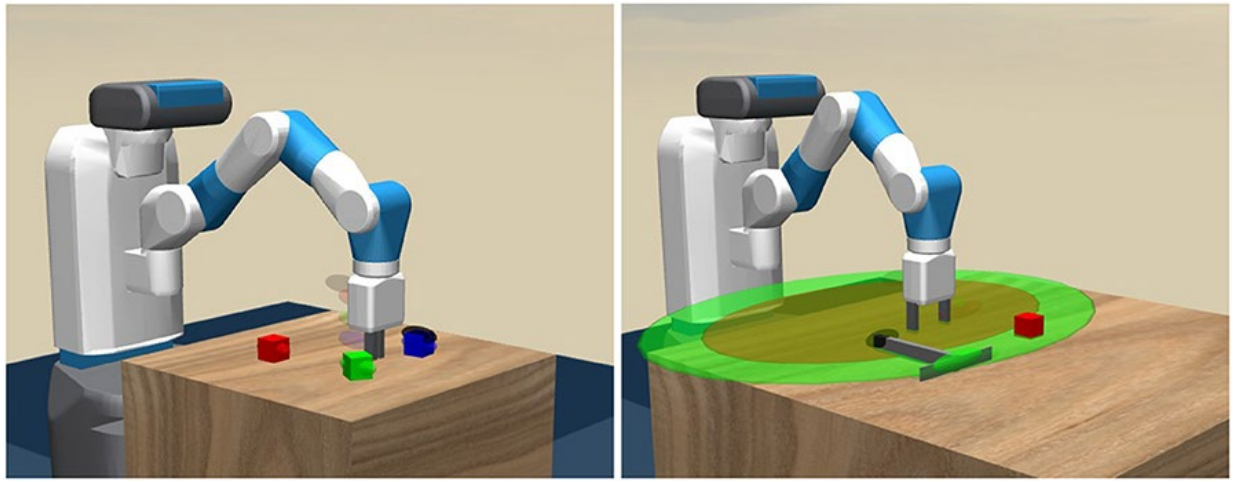
Additive game



Multiplicative game

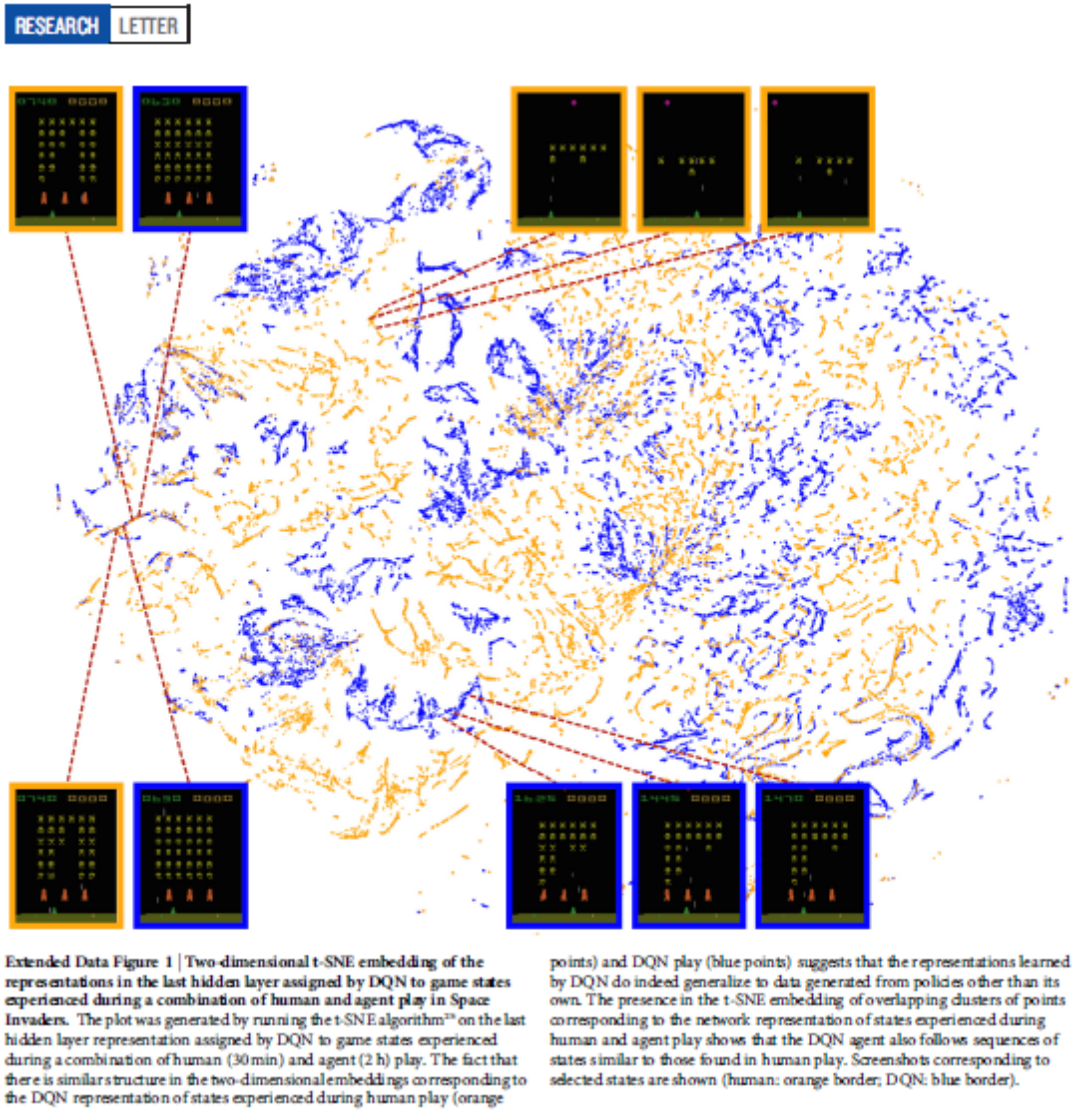


What is Reinforcement Learning (RL)



Eppe, M., Nguyen, P. D., & Wermter, S. (2019). From semantics to execution: Integrating action planning with reinforcement learning for robotic tool use. arXiv preprint arXiv:1905.09683.

Busa, V., (2018). *Open AI gym. Figure 2: FrozenLake-v0 board game: https://twice22.github.io/rl-part1/*



Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *nature*, *518*(7540), 529-533.

A brief introduction to Q-learning

Machine learning method:

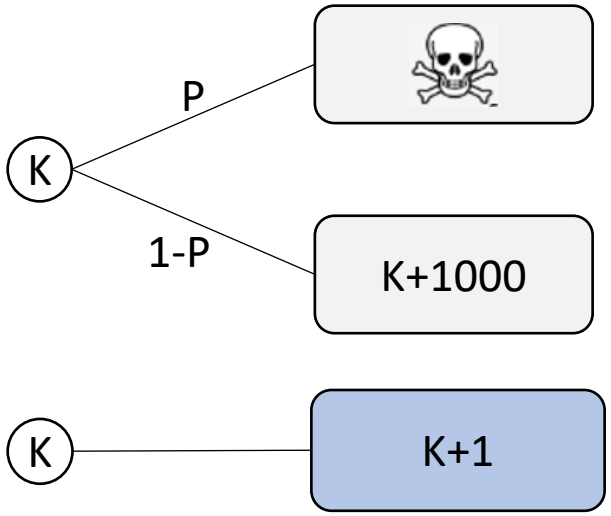
- Different from supervised or unsupervised learning
- Train agent from experience
- Basic Q-Learning

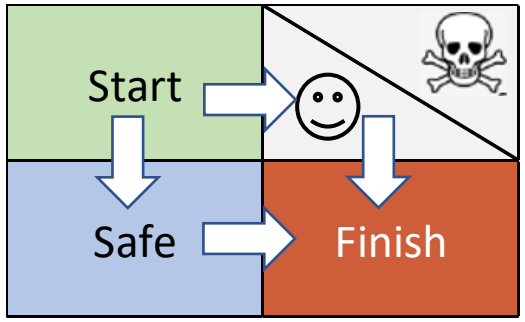
Initialize $Q(s,a), \forall s \in \mathcal{S}, a \in \mathcal{A}(s)$, arbitrarily, and $Q(\textit{terminal-state}, \cdot) = 0$
Repeat (for each episode):
  Initialize S
  Repeat (for each step of episode):
    Choose A from S using policy derived from Q (e.g., ϵ-greedy)
    Take action A, observe R, S'
    $Q(S,A) \leftarrow Q(S,A) + \alpha\big[R + \gamma \max_a Q(S',a) - Q(S,A)\big]$
    $S \leftarrow S'$;
  until S is terminal

Example (Frozen lake)

How do we translate the Casino-bets to the RL realm?

Additive bets

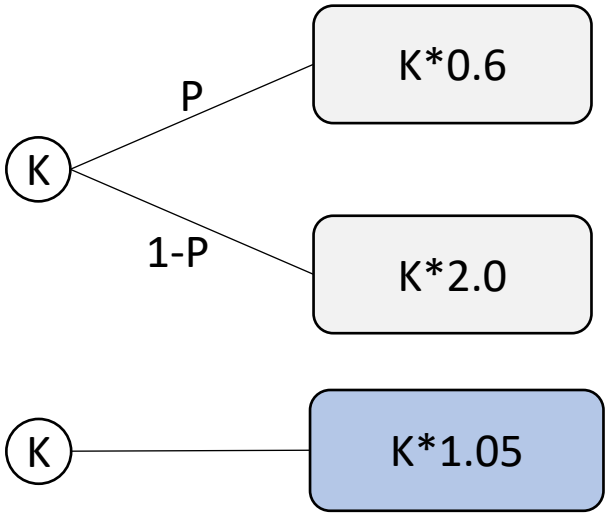


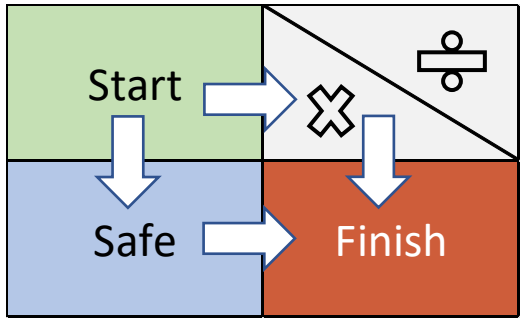


Casino bets

Reinforcement Learning

Multiplicative bets





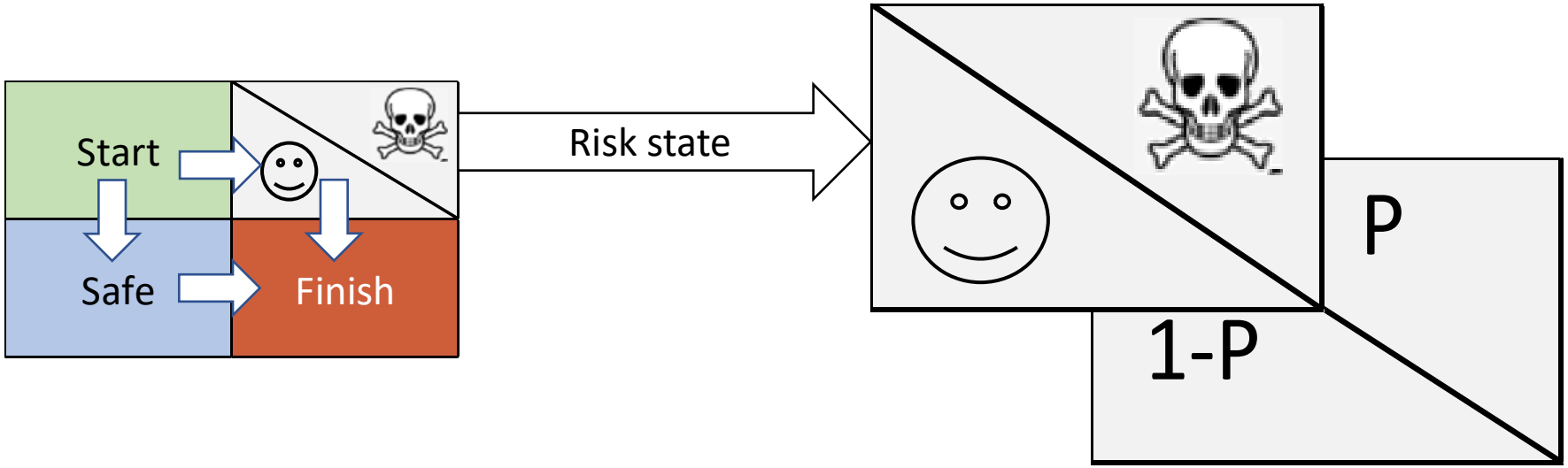
How to evaluate the RL decisions?

Location of inflection point:
(policy change)

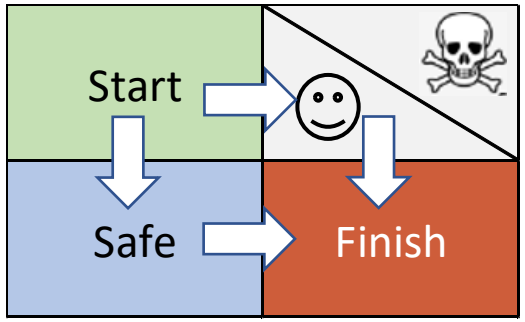
$$R_{safe} + R_{finish} = p * R_{fail} + (1-p) * (R_{risk} + R_{finish})$$
$$\Rightarrow p = \frac{R_{safe} - R_{risk}}{R_{fail} - R_{risk} - R_{finish}}$$

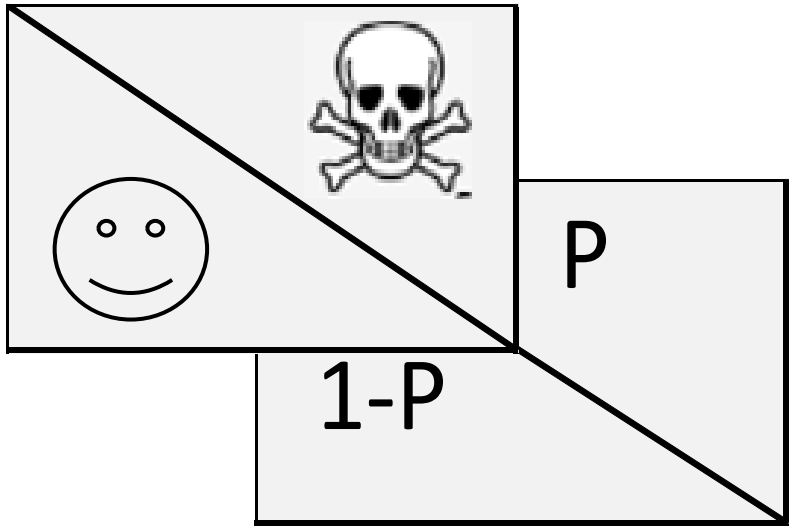
Slope tangent line at inflection point:
(rate of change)

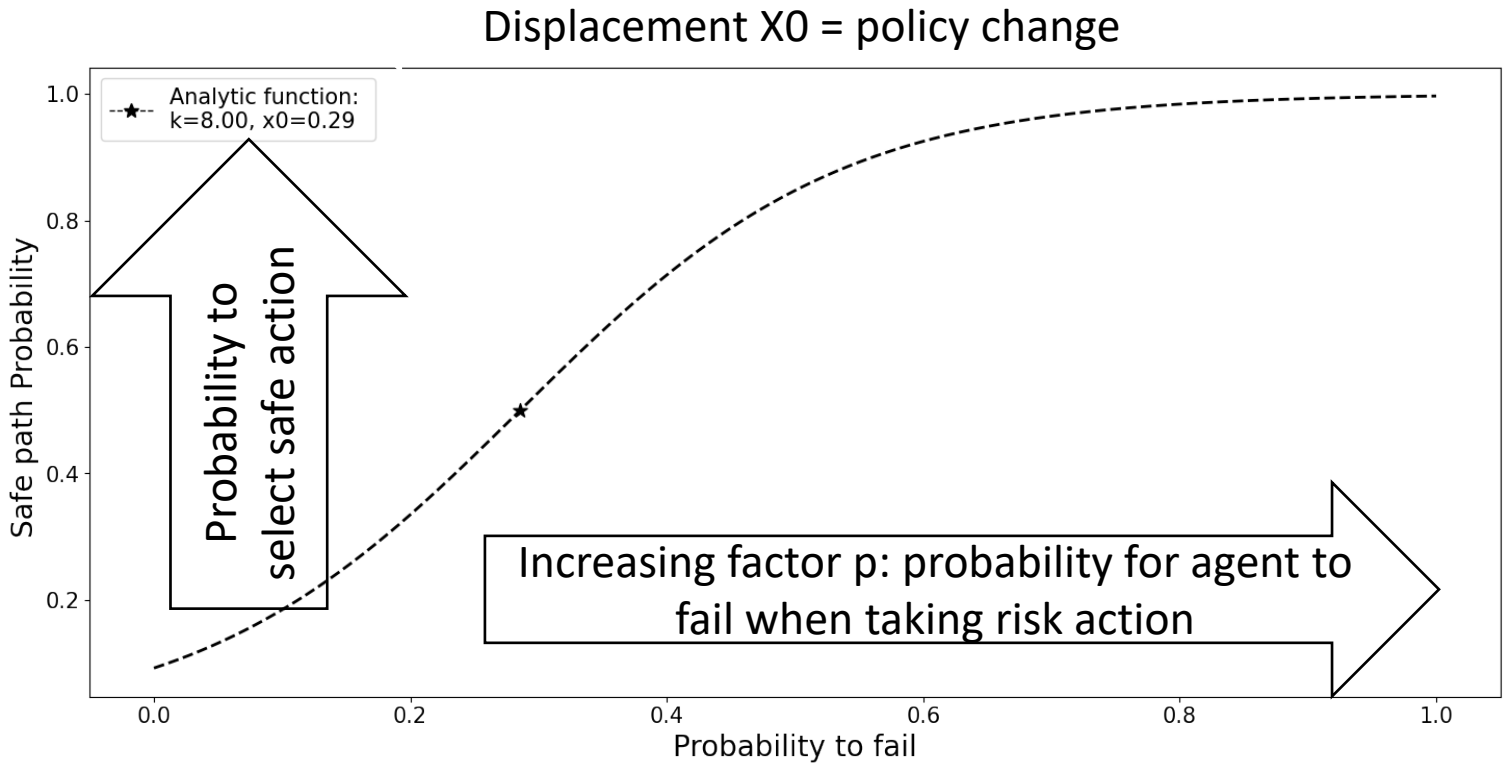
$$f(p) = R_{safe} - R_{risk} - p * (R_{fail}R_{risk} - R_{finish})$$
$$k = \frac{d}{dp} f(p)$$
$$k = R_{fail} - R_{risk} - R_{finish}$$



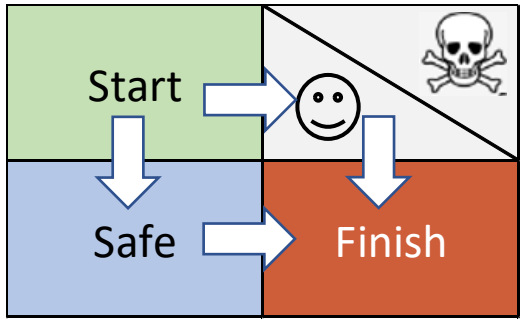
How do we evaluate a policy?

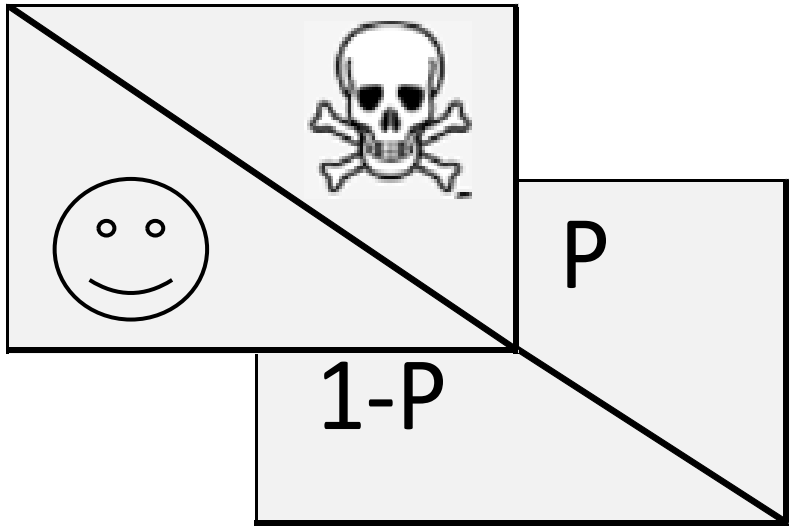


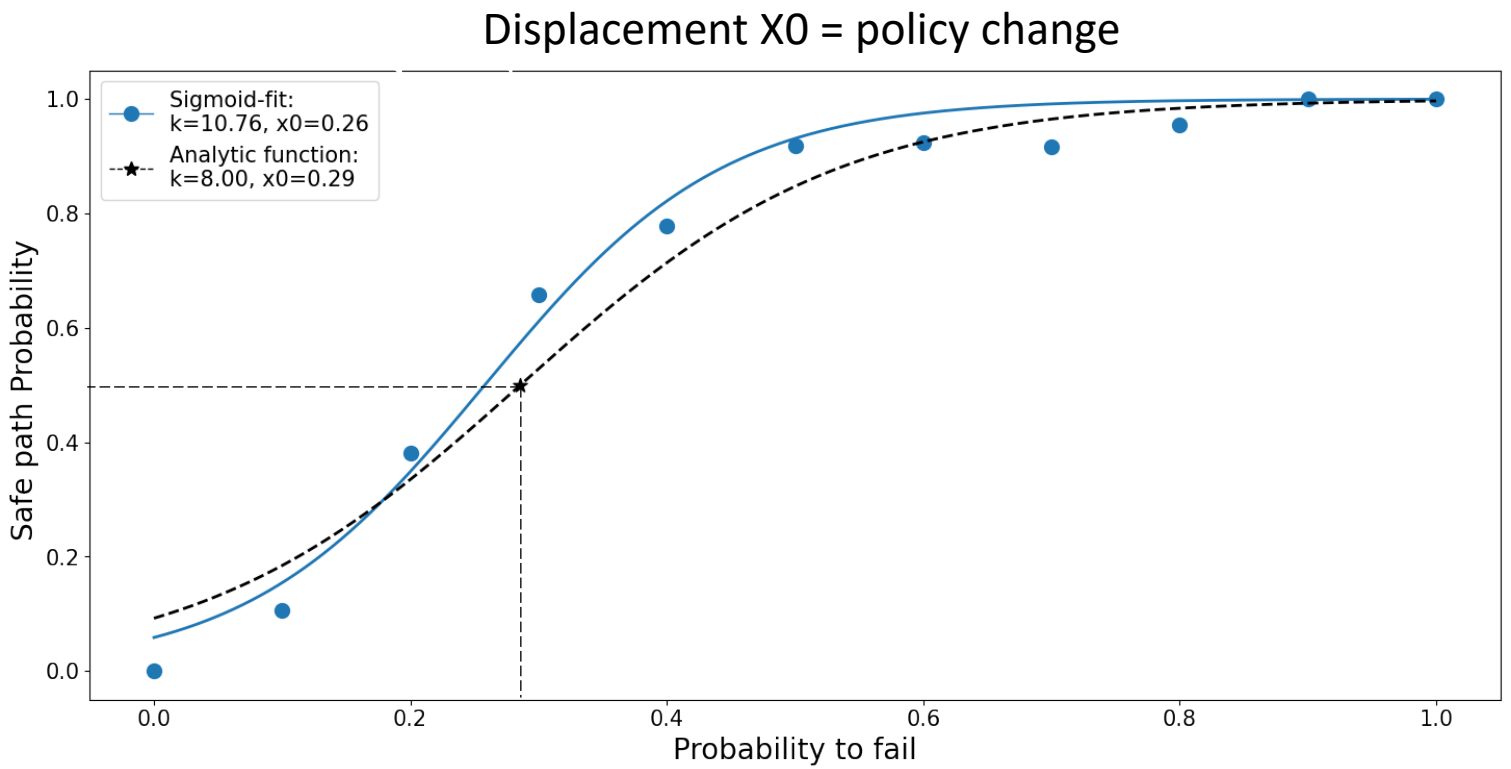




How do we evaluate a policy?



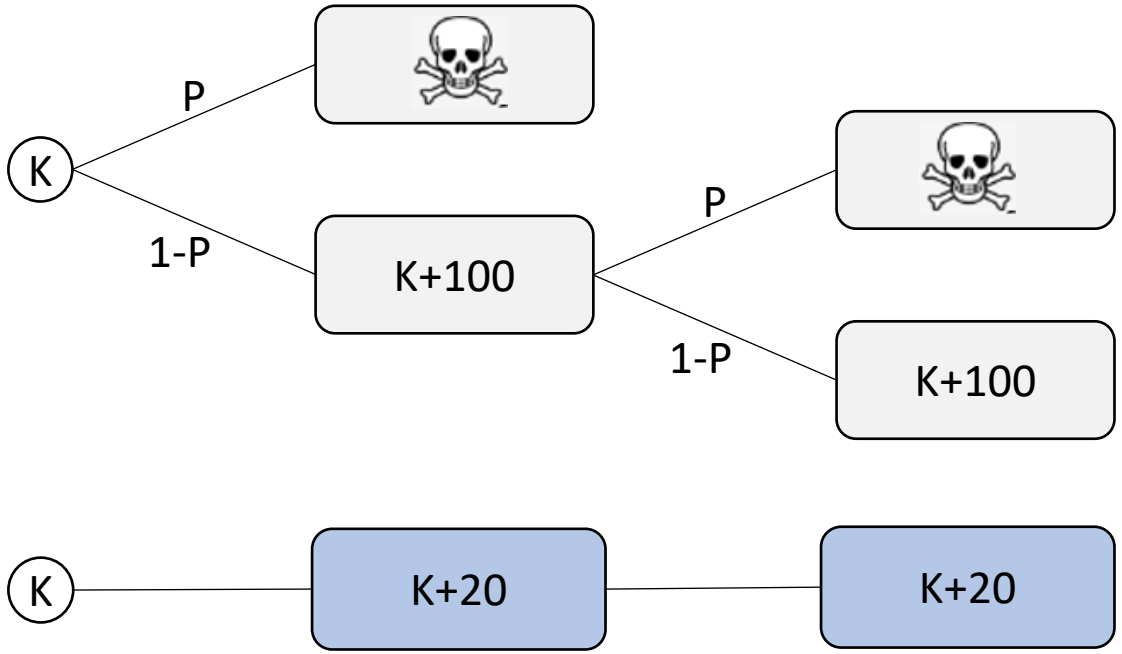




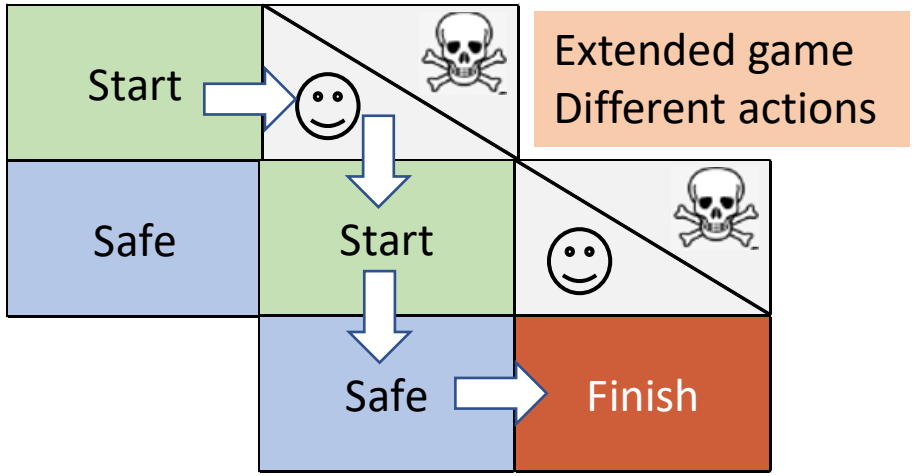
Introducing time-average for the agent: Repeated games before evaluation

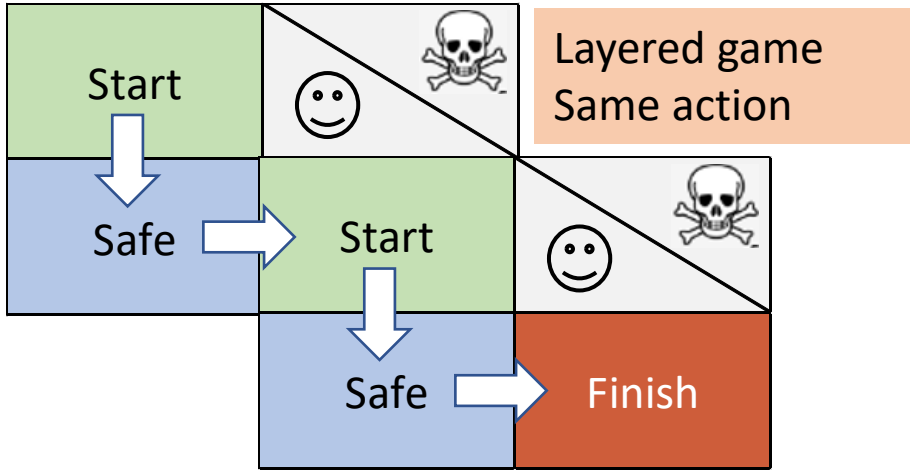
Real life:

Repeated bets

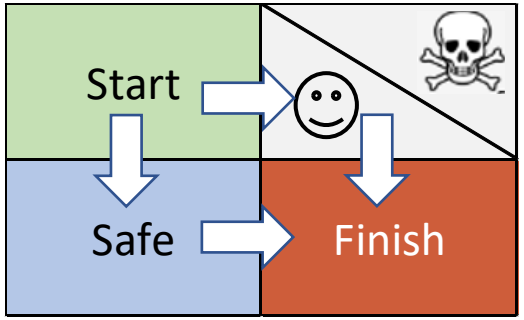


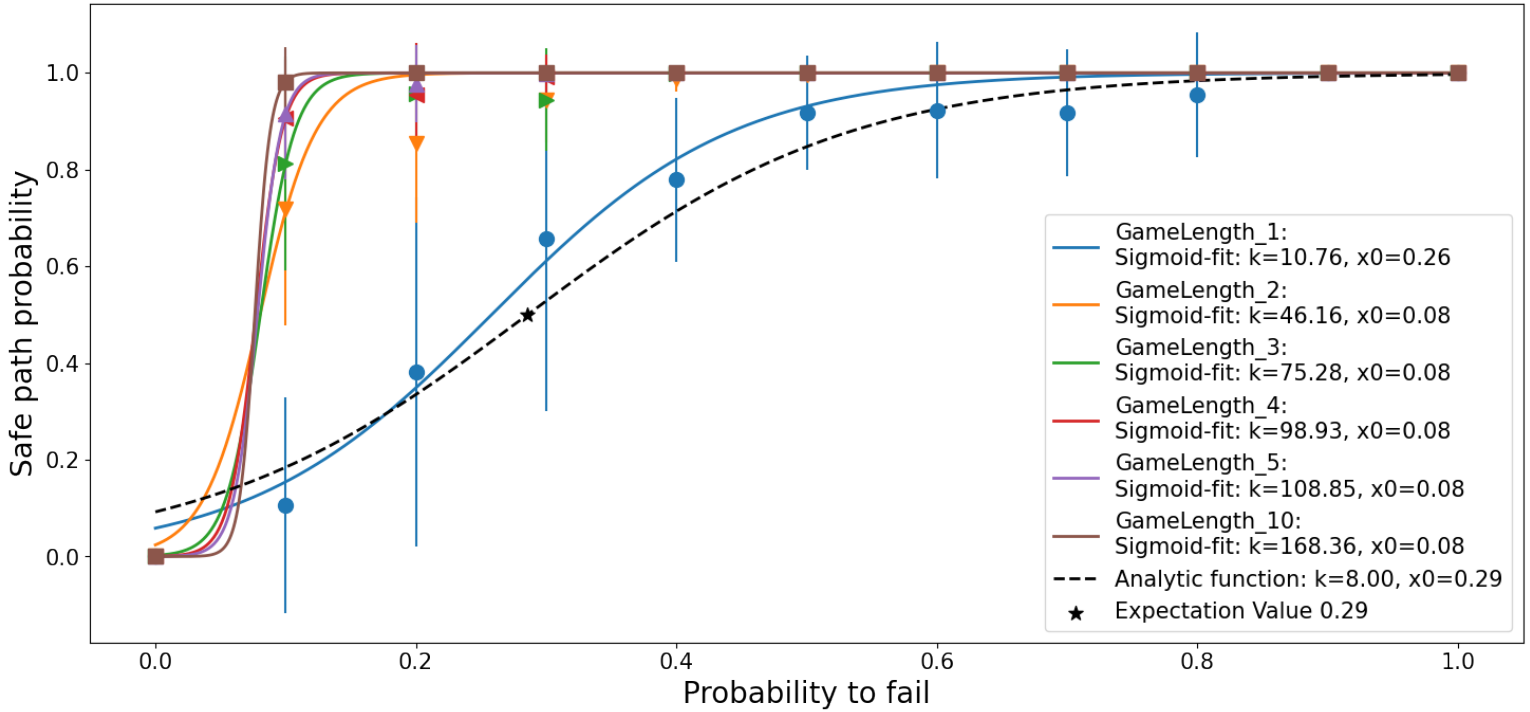
Reinforcement learning



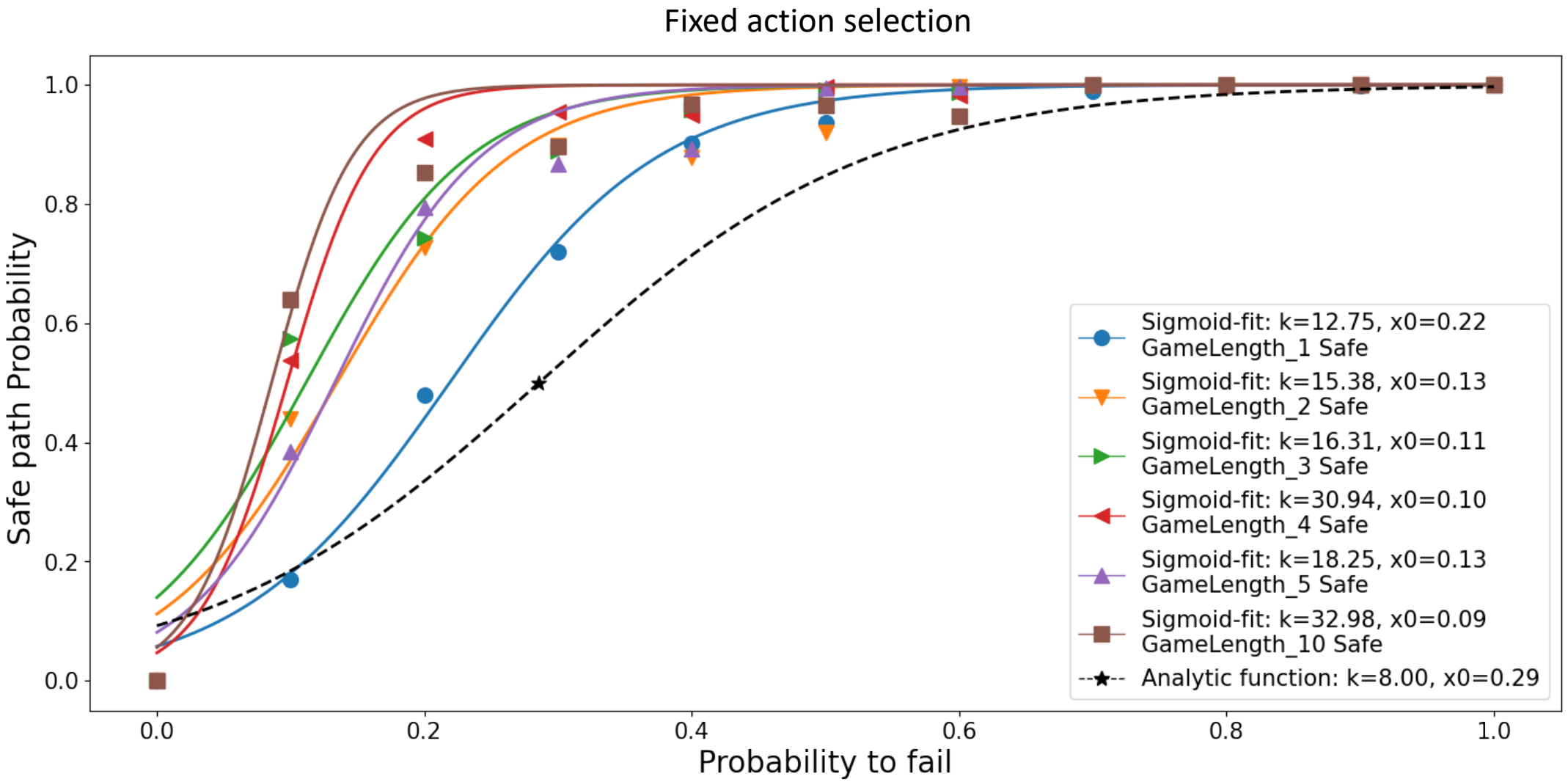


Additive models

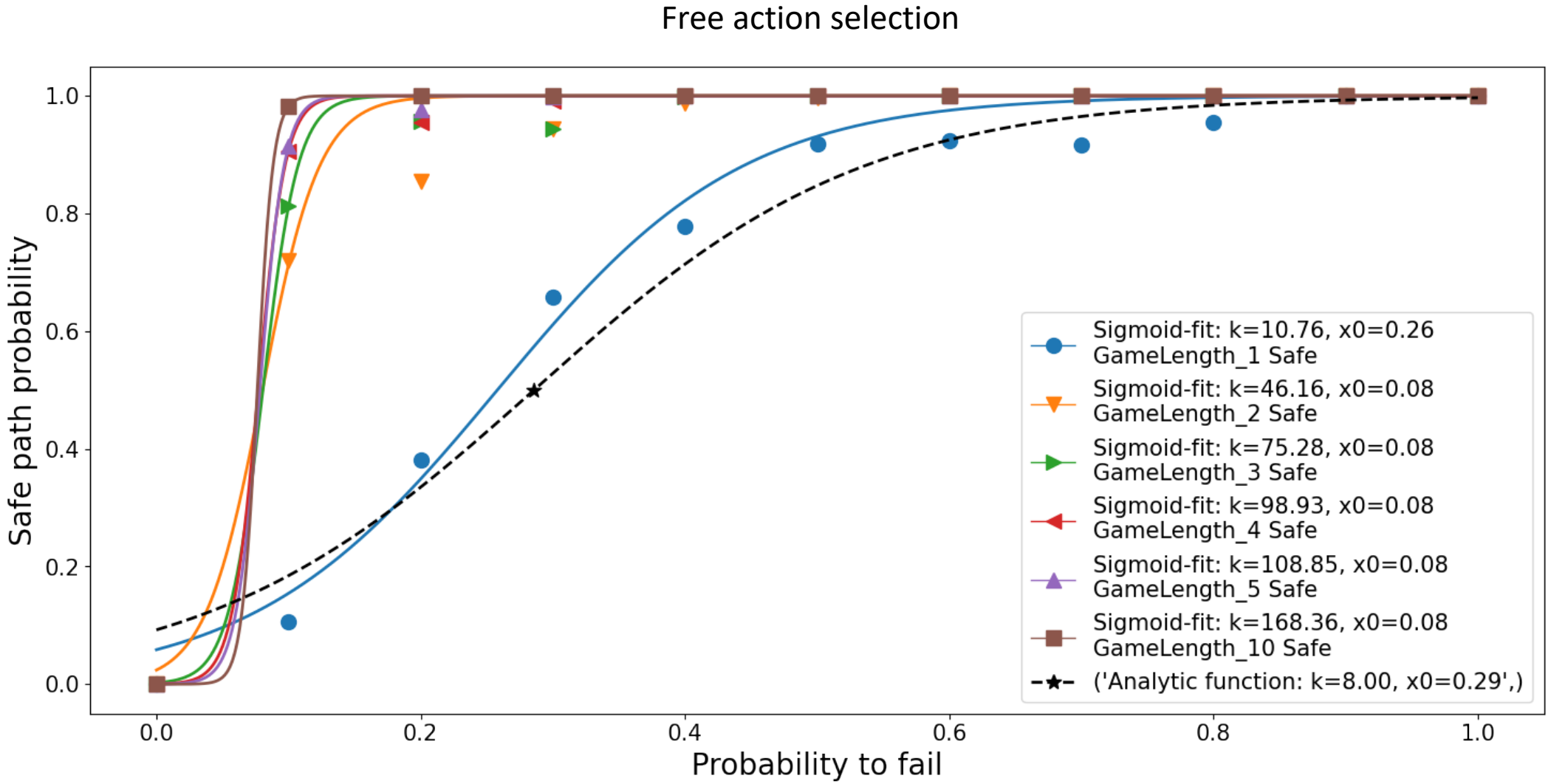




Additive model: Layered game

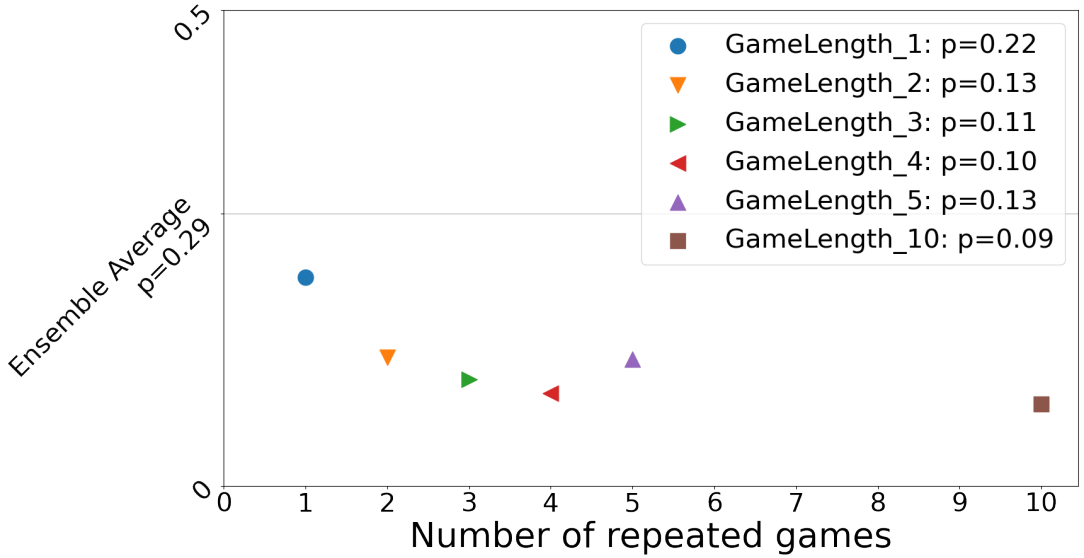


Additive model: Extended game

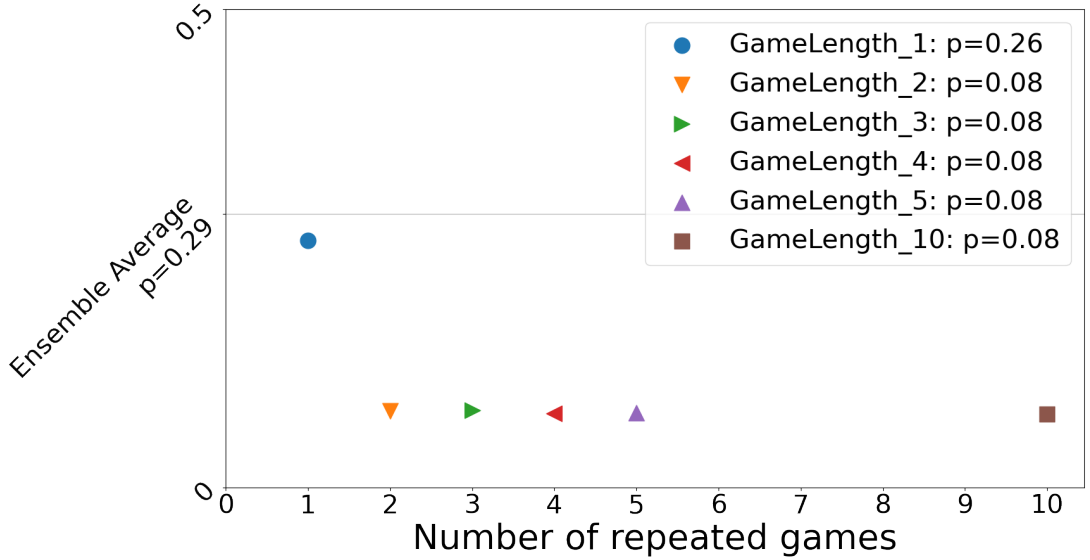


Summary: Additive models

Layered Model: Free actions



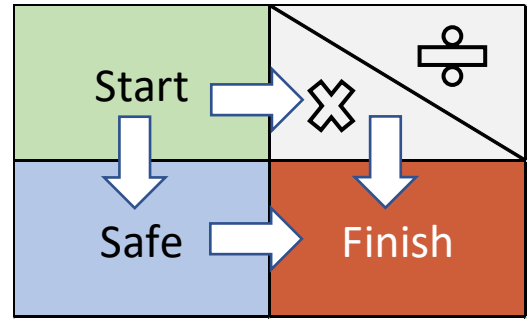
Extended Model: Fixed actions



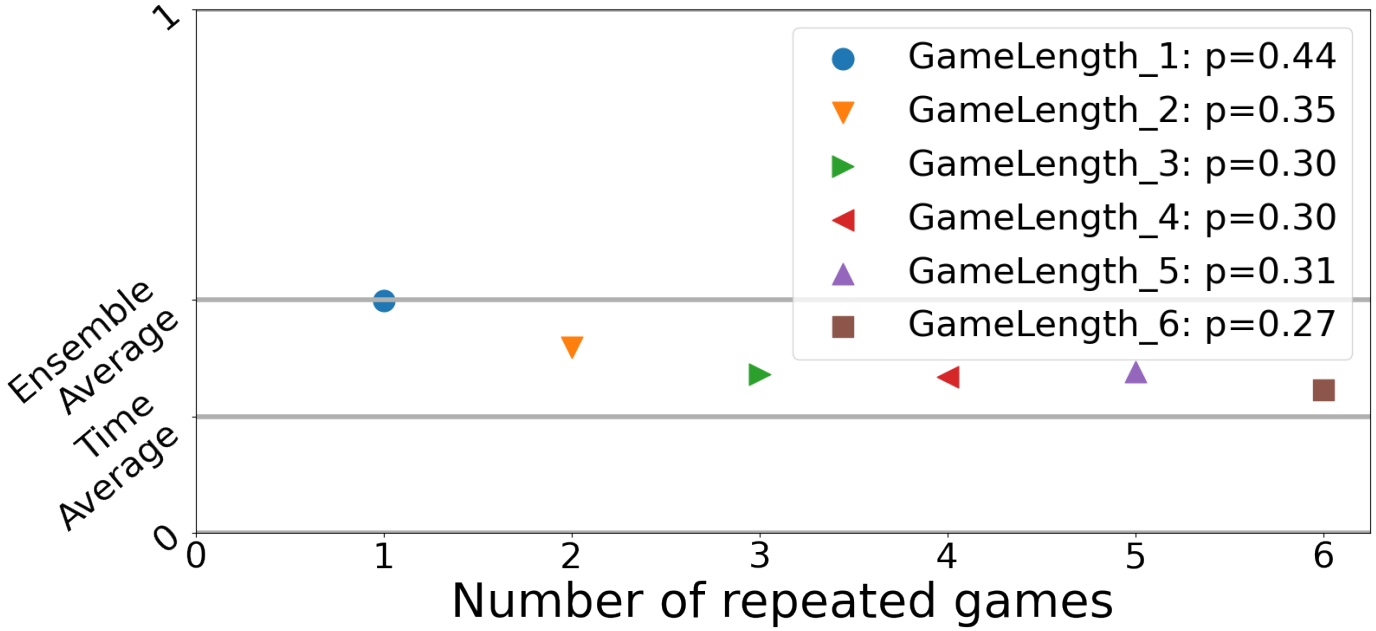
- Traditional RL algorithms evaluate the expected value of different choices and are thus unable to act optimally in non-ergodic contexts.
- We can let the algorithms experience ergodicity breaking by repeatedly asking the agent to play the games. That way, the policy also evolves in accordance with the optimal time-averaging.

Multiplicative model

Layered model



Multiplicative model: Conclusion



- Optimization to group average for one round
- Optimization to time average after modelling agent with repeated decisions

Conclusion

- Traditional RL algorithms evaluate the expected value of different choices and are thus unable to act optimally in non-ergodic contexts.
- We can let the algorithms experience ergodicity breaking by repeatedly asking the agent to play the games. That way, the policy also evolves in accordance with the optimal time-averaging.
- Results apply for two types of ergodicity breaking handled in the study. An additive model with ruin and a model with multiplicative dynamics


Thank you for listening.
Questions?