Nested Reasoning About Autonomous Agents Using Probabilistic Programs

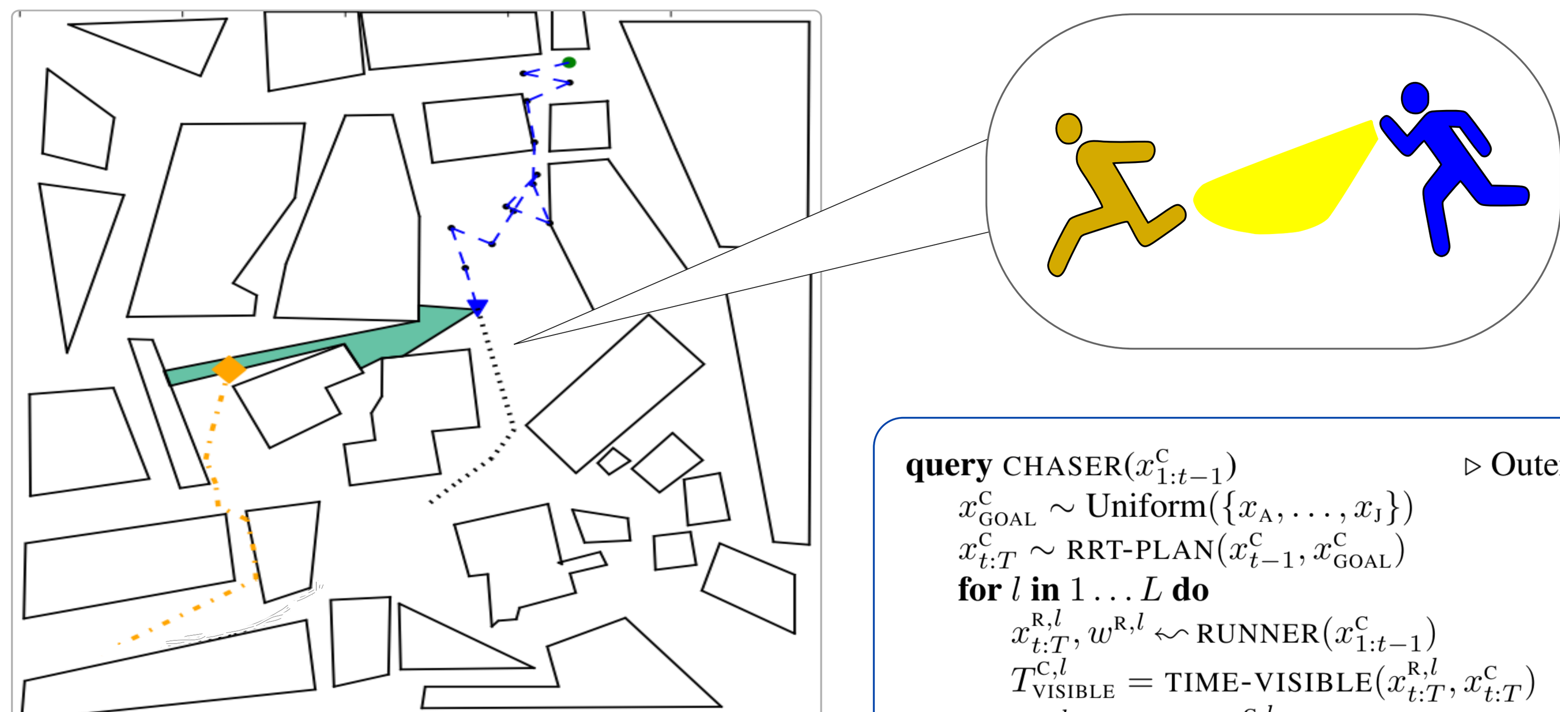
Iris R. Seaman
seaman.i@northeastern.edu

Jan-Willem van de Meent
j.vandemeent@northeastern.edu

David Wingate
wingated@cs.byu.edu

The Chaser-Runner Model

We explore how *theory of mind* can be implemented on high-uncertainty pursuit-evasion games using nested simulations in the form of probabilistic programs to reasoning about the plans of other agents.



Our probabilistic programs incorporate variety of complex primitives such as field-of-view calculations and path planners, which enable us to model quasi-realistic scenarios in a computationally tractable manner.

query EPISODE(x^{C}_{START}) ▷ Episode model
for k **in** $1 \ldots K$ **do**
  $x^{C,k}_{1} = x^{C}_{\text{START}}$
for t **in** $2 \ldots T$ **do**
  for k **in** $1 \ldots K$ **do**
    $x^{C,k}_{1:t}, w^{k}_{t} \leftsquigarrow \text{CHASER}(x^{C,k}_{1:t-1})$
  for k **in** $1 \ldots K$ **do**
    $a \sim \text{Categorical}\left(\frac{w^{1}_{t}}{\sum_k w^{k}_{t}}, \ldots, \frac{w^{K}_{t}}{\sum_k w^{k}_{t}}\right)$
    $x^{C,k}_{1:t}, w^{k}_{t} = x^{C,a}_{1:t}, \frac{1}{K}\sum_k w^{k}_{t}$
return $(x^{C,1}_{1:T}, w^{1}_{T}), \ldots, (x^{C,K}_{1:T}, w^{K}_{T})$

query CHASER($x^{C}_{1:t-1}$) ▷ Outer Model
  $x^{C}_{\text{GOAL}} \sim \text{Uniform}(\{x_A, \ldots, x_J\})$
  $x^{C}_{t:T} \sim \text{RRT-PLAN}(x^{C}_{t-1}, x^{C}_{\text{GOAL}})$
  for l **in** $1 \ldots L$ **do**
    $x^{R,l}_{1:T}, w^{R,l} \leftsquigarrow \text{RUNNER}(x^{C}_{1:t-1})$
    $T^{C,l}_{\text{VISIBLE}} = \text{TIME-VISIBLE}(x^{R,l}_{t:T}, x^{C}_{t:T})$
    $w^{R,l} = \exp(\alpha\, T^{C,l}_{\text{VISIBLE}})$
  return $x^{C}_{1:t}, w^{C} \cdot \left(\frac{1}{L}\sum_l w^{R,l}\right)$

query RUNNER($x^{C}_{1:t-1}$) ▷ Middle Model
  $x^{R}_{\text{START}} \sim \text{Uniform}(\{x_A, \ldots, x_J\})$
  $x^{R}_{\text{GOAL}} \sim \text{Uniform}(\{x_A, \ldots, x_J\})$
  $x^{R}_{1:T} \sim \text{RRT-PLAN}(x^{R}_{\text{START}}, x^{R}_{\text{GOAL}})$
  $\tilde{x}^{C}_{t:T}, \tilde{w}^{C} \leftsquigarrow \text{NAIVE-CHASER}(x^{C}_{t-1})$
  $T^{R}_{\text{VISIBLE}} = \text{TIME-VISIBLE}(x^{R}_{1:T}, \{x^{C}_{1:t-1}, \tilde{x}^{C}_{t:T}\})$
  $w^{R} = \exp(-\alpha\, T^{R}_{\text{VISIBLE}})$
  return $x^{R}_{1:T}, w^{R} \cdot \tilde{w}^{C}$

query NAIVE-CHASER(x^{C}_{t-1}) ▷ Inner Model
  $\tilde{x}^{C}_{\text{GOAL}} \sim \text{Uniform}(\{x_A, \ldots, x_J\})$
  $\tilde{x}^{C}_{t:T} \sim \text{RRT-PLAN}(x^{C}_{t-1}, \tilde{x}^{C}_{\text{GOAL}})$
  return $\tilde{x}^{C}_{t:T}, 1$

Planning as Inference

$$\pi(x) = \gamma(x)/Z$$

$$\gamma(x) = \exp(R(x))p(x)$$

$$Z = \mathbb{E}[\exp(R(x))]$$

$$\gamma_t(x^{C}_{t:T}, \tilde{x}^{C}_{t:T}, x^{R}_{1:T} \mid x^{R}_{t-1}) = \exp[\alpha(T^{C}_{\text{VIS}} - T^{R}_{\text{VIS}})]\; p(x^{C}_{t:T}|x^{C}_{t-1})\; p(\tilde{x}^{C}_{t:T}|x^{C}_{t-1})\; p(x^{C}_{1:T})$$

Nested Importance Sampling $\quad w^{k}_{t} = \frac{1}{L}\sum_{l=1}^{L} \exp\left[R(x^{C,k}_{1:T}, x^{R,k,l}_{1:T})\right]$

Nested Conditioning

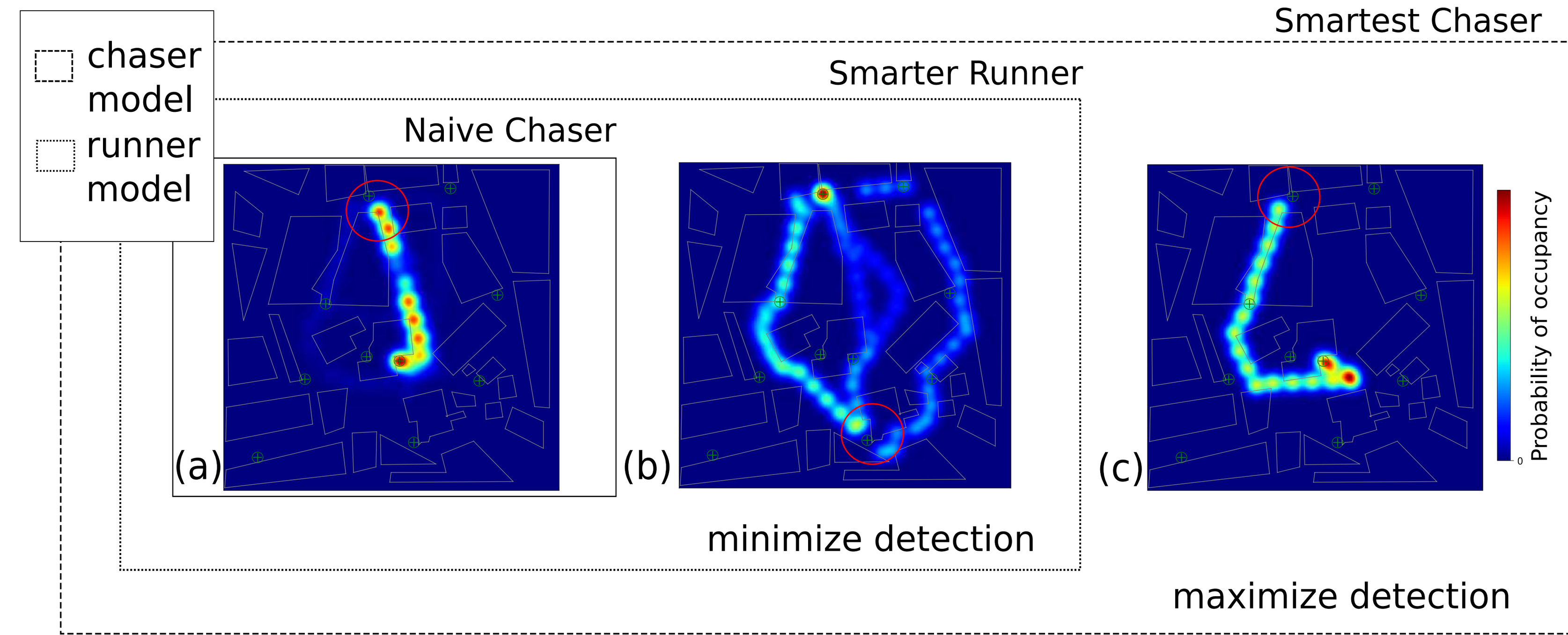
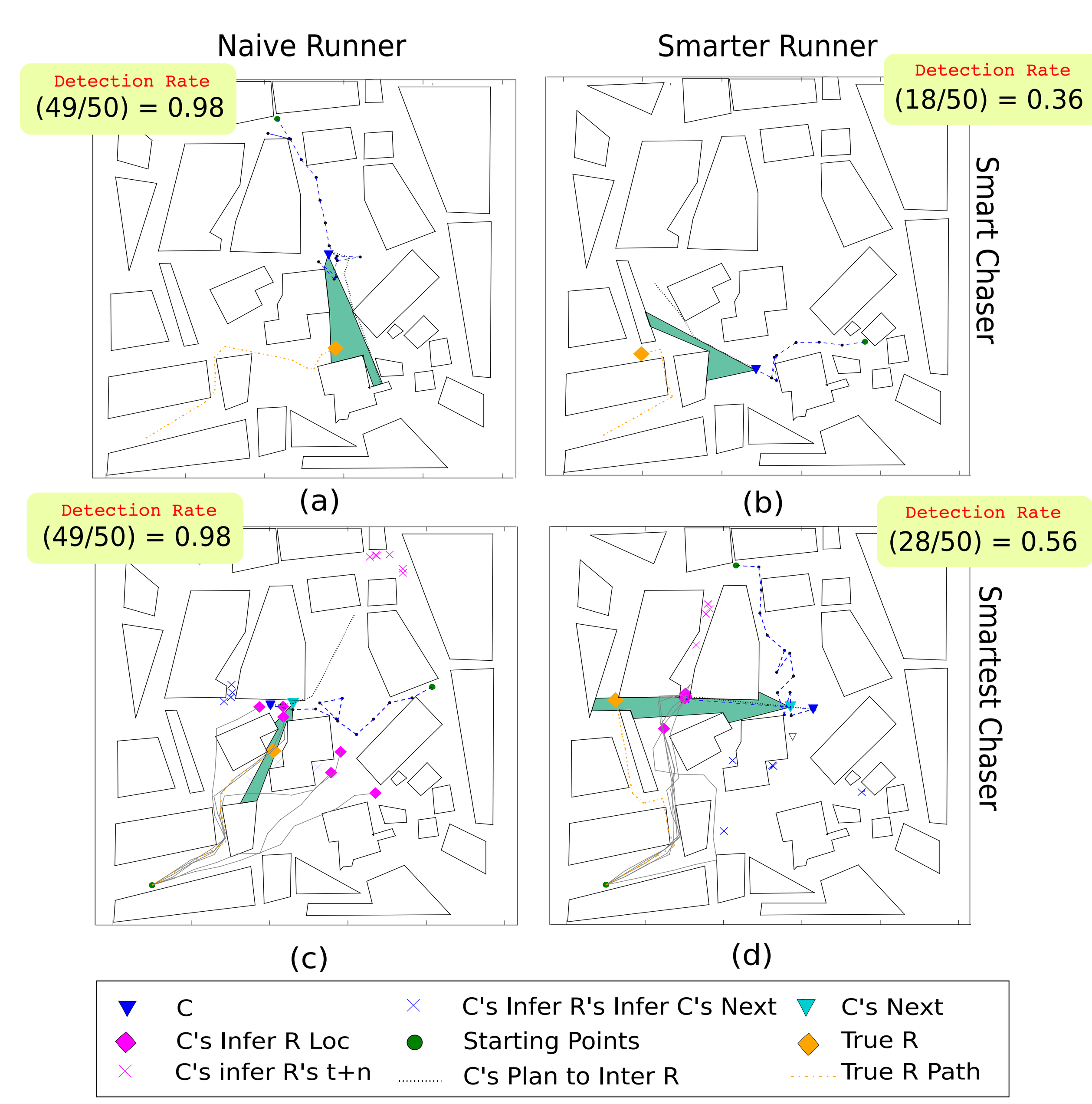


Figure 1 Chaser and runner trajectories in the innermost, middlemost, and outermost models where locations circled in red are the starting locations for each agent. We show posterior distributions of *L runner* and *naive chaser* paths, when *(K,L)* = 128, 16 for a single resampled sample *k*. We condition the start and goal locations in this experiment.

Detection Experiments



We compare detection rates in the full chaser-runnermodel to detection rates in three simplified models

Discussion These 4 scenarios illustrate that when the runner reasons more deeply, he evades more effectively: conversly when the chaser reasons more deeply, he intercepts more effectively.

Futhermore, we show that a single, unified inference algorithm can uncover a wide variety of intuitive, rational behaviors or both the runner and the chaser.

Figure 2 We compare detection rates in the full chaser-runner model to detection rate in three simplified models. Average detection rate over 50 restarts for each scenario.

Dependence on Sample Budget

Figure 3 Log mean log weights, $\log \bar{Z}$, and Fractional ESS as a function of time for each sample budget.

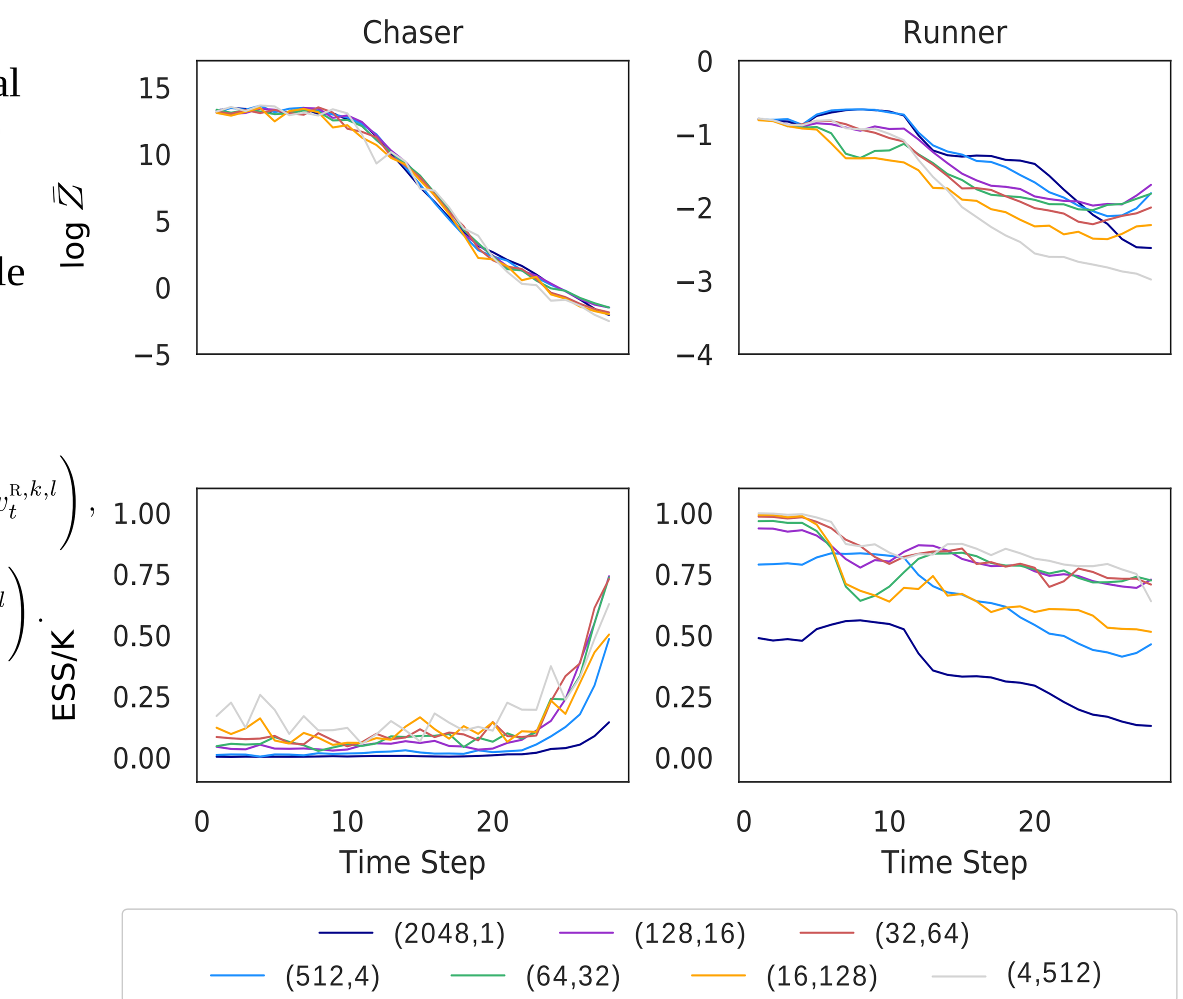
Top Row $\log \bar{Z}^{R}$ for the middle most model (left), and $\log \bar{Z}^{C}$ the outermost model, (right).

$$\log \bar{Z}^{C}_{t} = \frac{1}{R}\sum_{r=1}^{R} \log\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{l=1}^{L} w^{C,k,l}_{t} w^{R,k,l}_{t}\right),$$

$$\log \bar{Z}^{R}_{t} = \frac{1}{R}\sum_{r=1}^{R} \log\left(\frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L} w^{R,k,l}_{t}\right).$$

Bottom Row The fractional ESS for each varying K and L.

$$\text{ESS} = \frac{\left(\sum_{k=1}^{K}\sum_{l=1}^{L} w^{k,l}\right)^2}{\left(\sum_{k=1}^{K}\sum_{l=1}^{L} (w^{k,l})^2\right)}$$



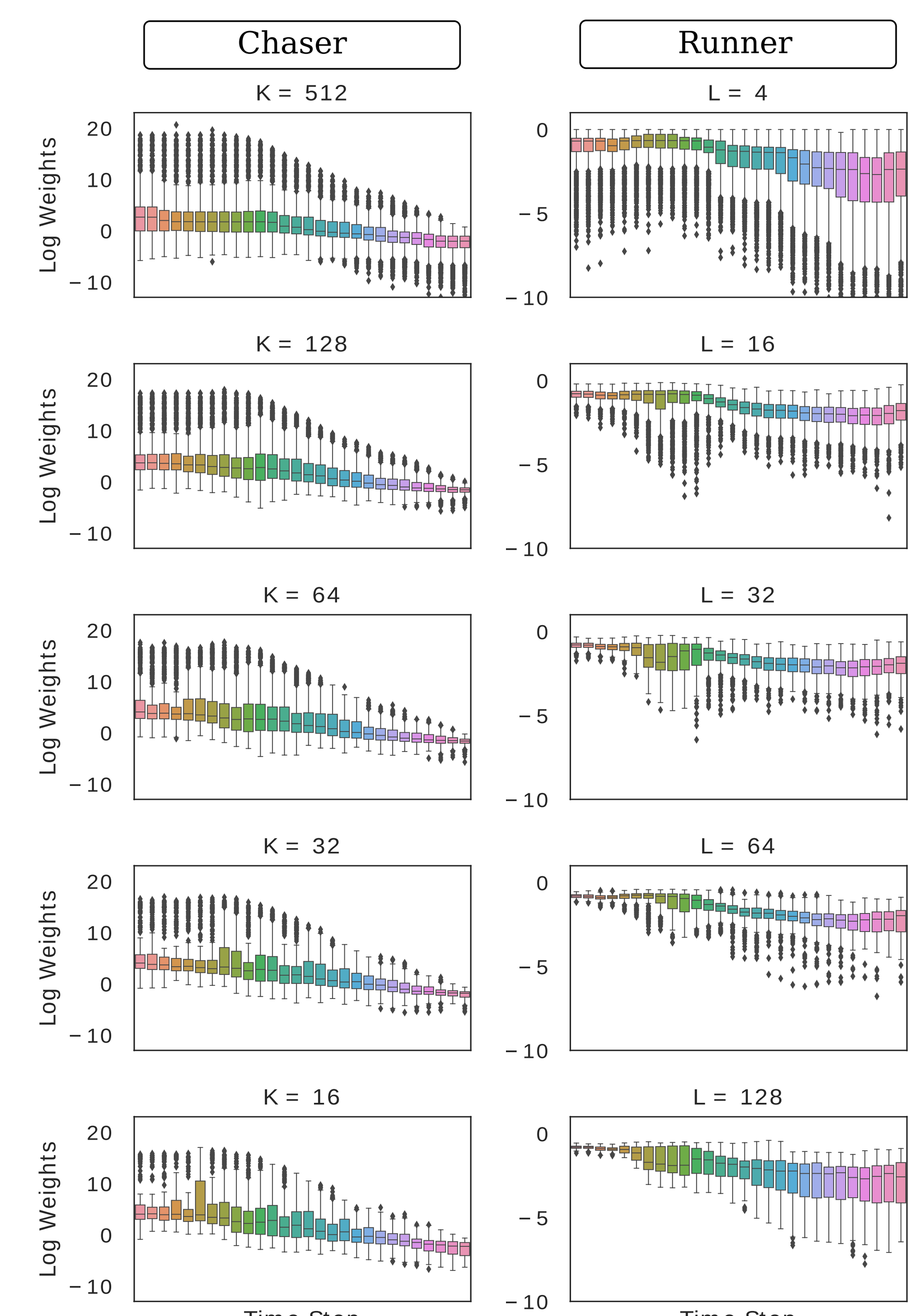


Figure 4 Box plots showing quantiles with respect to restarts for log mean weights for varying K and L.

We show higher median log weights and less outliers as K decreases and L increases, (K, L) = (16,128) results show that the log weights are less robust when we draw less K samples.

Conclusion We have introduced a high-uncertainty variant of pursuit-evasion games where agents are required to reason about other agents' reasoning in order to accomplish their respective goals.

We empirically demonstrate that nested Bayesian reasoning leads to rational behaviors, incorporating theory of mind outperforms non-nested models, and nested reasoning results in lower-variance estimates of expected utility.

https://arxiv.org/abs/1812.01569