Human Visual Search as a Deep Reinforcement Learning Solution to a POMDP

Aditya Acharya [1] Xiuli Chen [1] Christopher W. Myers [2] Richard L. Lewis [3] Andrew Howes [1]

[1]University of Birmingham

[2]Air Force Research Laboratory

[3]University of Michigan,Ann Arbor, USA

Introduction

Visual Search

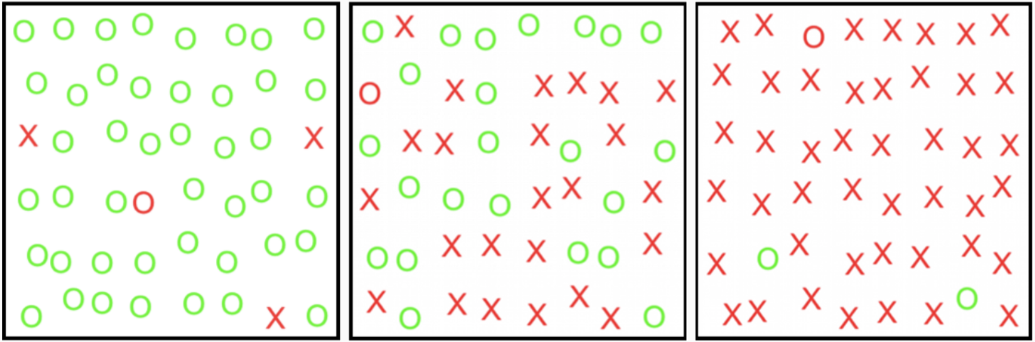
- In a visual search task people use a series of eye movements and fixations to find a desired target.
- In a typical laboratory visual search task, participants are asked to find a visual target amongst distractors.
- the number and similarity of distractors can be varied.

Overview

- Introduction to the Distractor-Ratio Task [4].
- Walk-through of a new model that adapts to human information processing constraints, task ecology and utility/reward.
- The model is based on a Partially Observable Markov Decision Process (POMDP).
- … and uses a Deep Q-Network (DQN) to find computationally rational strategies.
- Present model performance and compare it with a heuristic model.

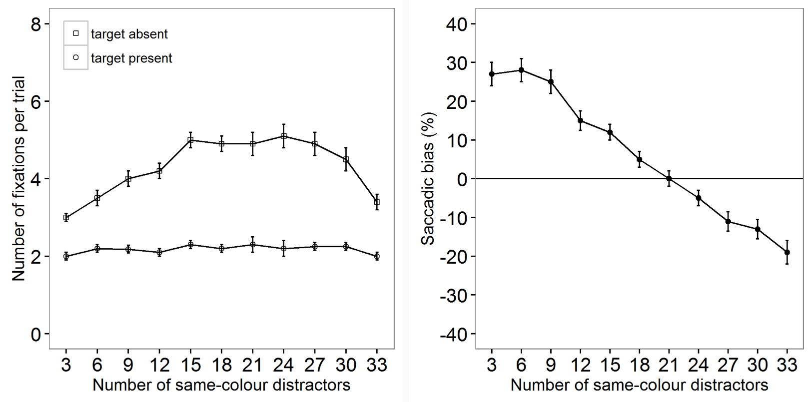
The Task

Distractor-Ratio



(a) (b) (c)

Figure 1: Distractor ratio stimuli with ratio distributions: (a) 3:45, (b) 24:24, (c) 46:2 and target stimuli: red coloured letter O.



(a) (b)

Figure 2: (a) Average number of fixations per trial as a function of the number of distractors sharing colour with the search target. (b) Saccadic bias (the difference between the observed frequency and chance performance) as a function of the number of same-colour distractors [4].

The Model

Framework

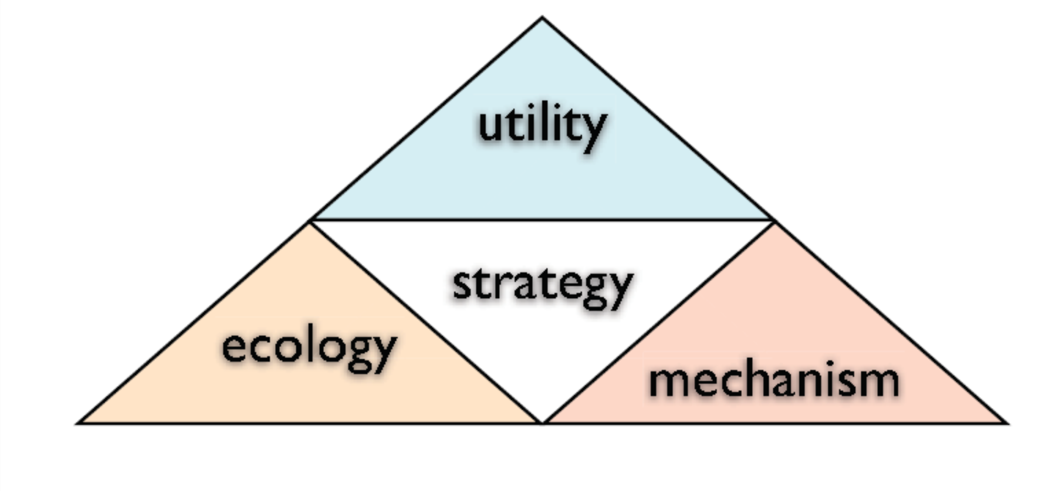
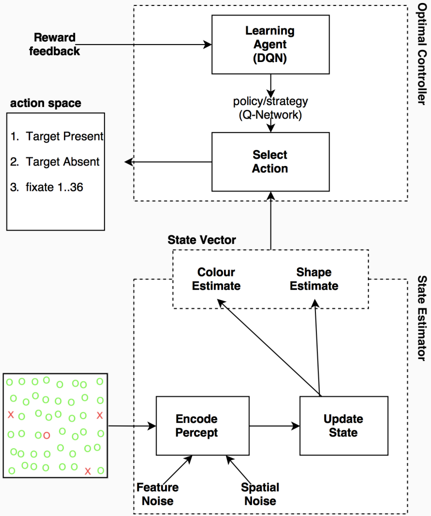


Figure 3: Adaptive Interaction Framework [3]

Framework

- We frame the visual search task as a solution to **POMDP**.
- A POMDP is defined by the following:
  - Set of states S, set of actions A, set of observations O
  - Transition model T(s, a, s')
  - Reward model R(s)
- Probability distribution over states, i.e., Belief State is maintained. Since, states are not directly observable.

Framework



Reward feedback
Learning Agent (DQN)
policy/strategy (Q-Network)
Select Action
Optimal Controller
action space
1. Target Present
2. Target Absent
3. fixate 1..36
State Vector
Colour Estimate
Shape Estimate
State Estimator
Encode Percept
Update State
Feature Noise
Spatial Noise

Observation Model

Two sources of uncertainty is encoded in the model.

Feature Noise: The human eye's ability to discriminate and perceive objects degrades with eccentricity according to a hyperbolic function.

Spatial Noise: Information in parafovea erroneously combine features from one location with adjacent locations.

Both sources of noise have been shown to be essential to modeling the DR Effect [2].

State Estimation

- The colour and shape observations are integrated across fixations, using naive Bayesian inference.
- These colour and shape estimates are then combined by element-wise multiplication to give a combined representation.

Reward Function

The reward distribution was defined as follows:

Reward	Action
+10	for correct response
-10	for incorrect response
-1	for fixation on a location

The penalty on each fixation imposes a speed-accuracy trade-off.

Heuristic Model

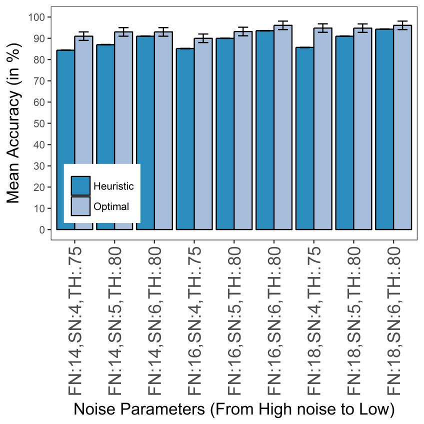
- An **alternative** to the (computationally rational) POMDP model.
- Utilizes the same state estimation.
- ‘Look for Best’ Heuristic strategy.
  - uses a MAP-like strategy to determine where to fixate next.
  - uses a thresholded stopping rule.

Results

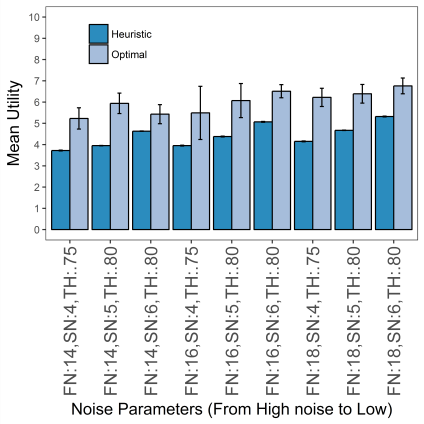
Results

Model simulations comparing the POMDP model and the heuristic model against human data

Accuracy and Utility



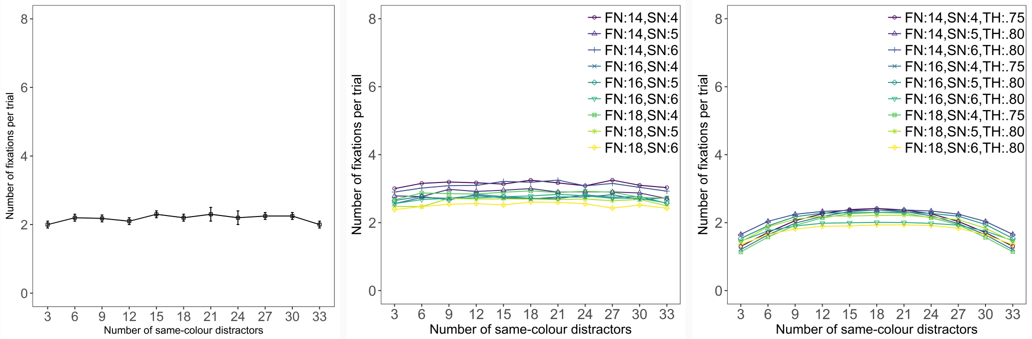
(a)



(b)

Figure 4: (a) Mean accuracy achieved by both models plotted against different noise parameter settings. (b) Mean utility gained by both models plotted against different noise parameter settings.

Distractor-Ratio for Target Present



(a) (b) R^2 = 0.98 (c) R^2 = 0.95

Figure 5: Number of fixations as a function of same-colour distractors for (a) Human (b) the Optimal Control model, (c) the Heuristic model when target is present.

Distractor-Ratio for Target Absent

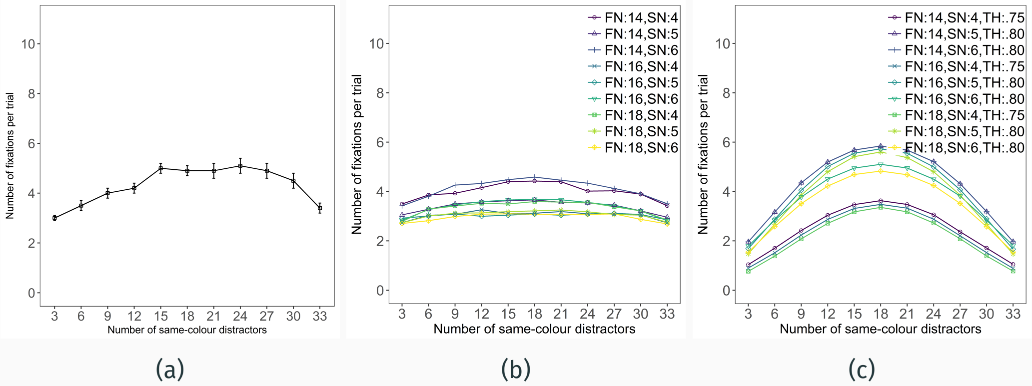
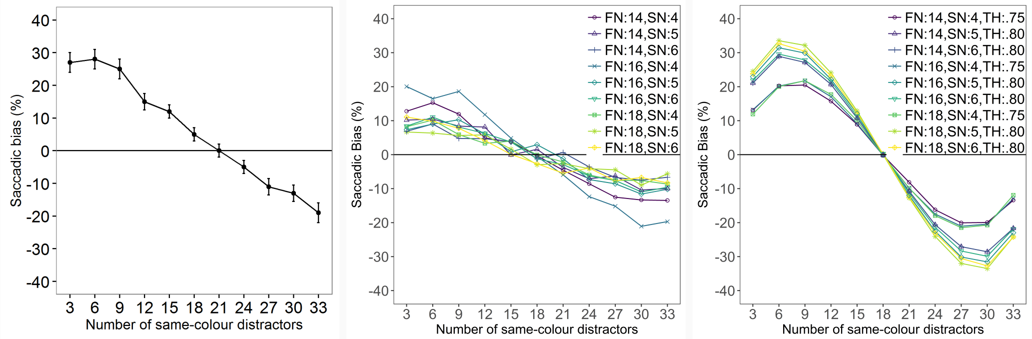


Figure 6: Number of fixations as a function of same-colour distractors for (a) Human (b) the Optimal Control model, (c) the Heuristic model when target is absent.

Saccadic Selectivity for Target Absent



(a) (b) R^2 = 0.97 (c) R^2 = 0.94

Figure 7: Saccadic bias as a function of the number of same-colour distractors for (a) Human (b) the Optimal Control model, (c) the Heuristic model when target is absent.

Conclusion

Summary

- Showed the distractor-ratio effect is the consequence of an approximately optimal adaptation to the constraints imposed by the human visual information processing system.
- Application of POMDP framework for framing of the distractor-ratio problem.
- Application of Deep Q-Learning [1] to determine the approximately optimal policy given a theory of human visual information processing capacities.

Future Work

- Extend architecture to incorporate recurrent network for an end-to-end learning.
- Explore parameter space to better fit human data.

Questions?

References i

V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al.
Human-level control through deep reinforcement learning.
Nature, 518(7540):529–533, 2015.

C. W. Myers, R. L. Lewis, and A. Howes.
Bounded optimal state estimation and control in visual search: Explaining distractor ratio effects.
In *CogSci*, 2013.

S. J. Payne and A. Howes.
Adaptive interaction: A utility maximization approach to understanding human interaction with technology.
Synthesis Lectures on Human-Centered Informatics, 6(1):1–111, 2013.

J. Shen, E. M. Reingold, and M. Pomplun.
Guidance of eye movements during conjunctive visual search: the distractor-ratio effect.
Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 57(2):76, 2003.