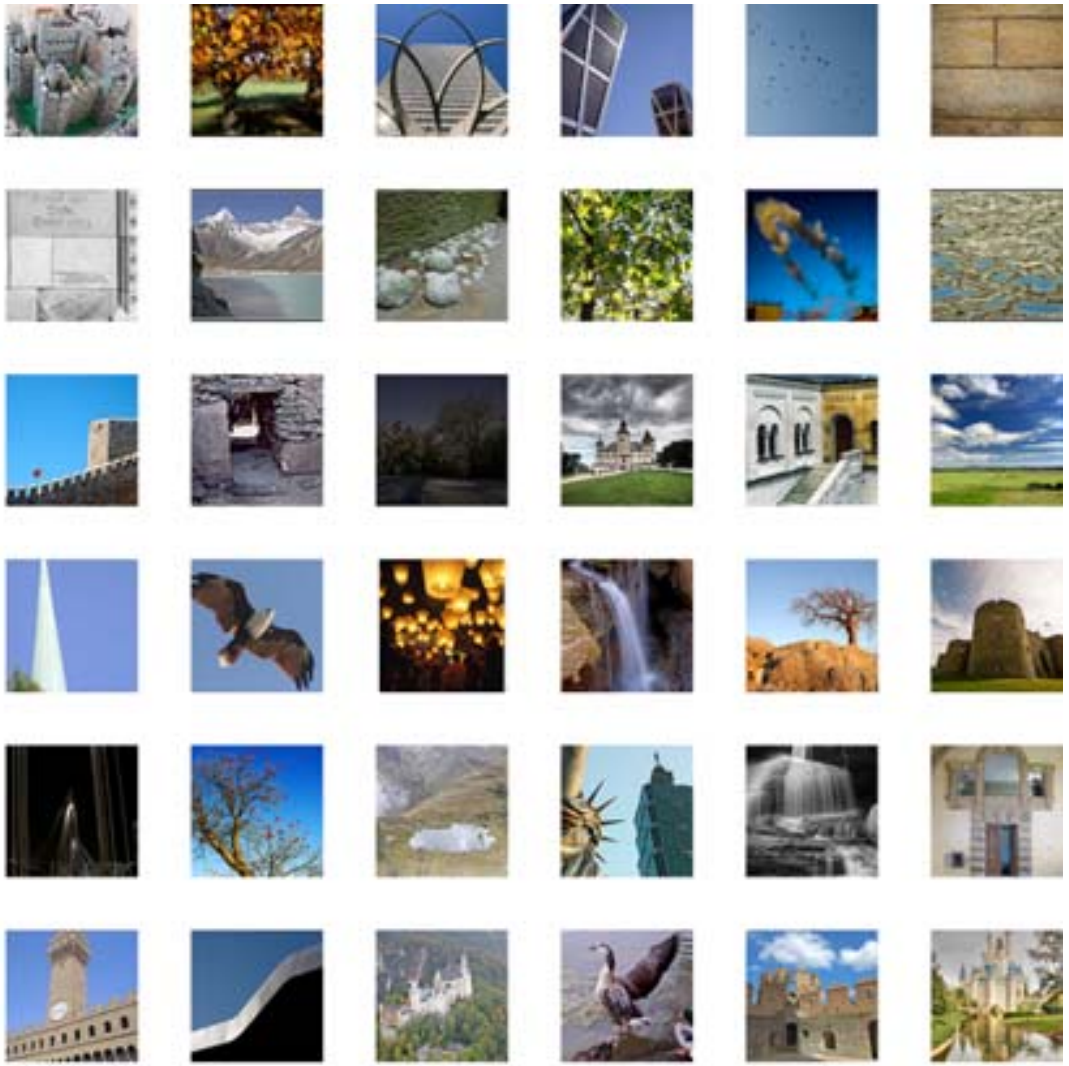
Visual Search

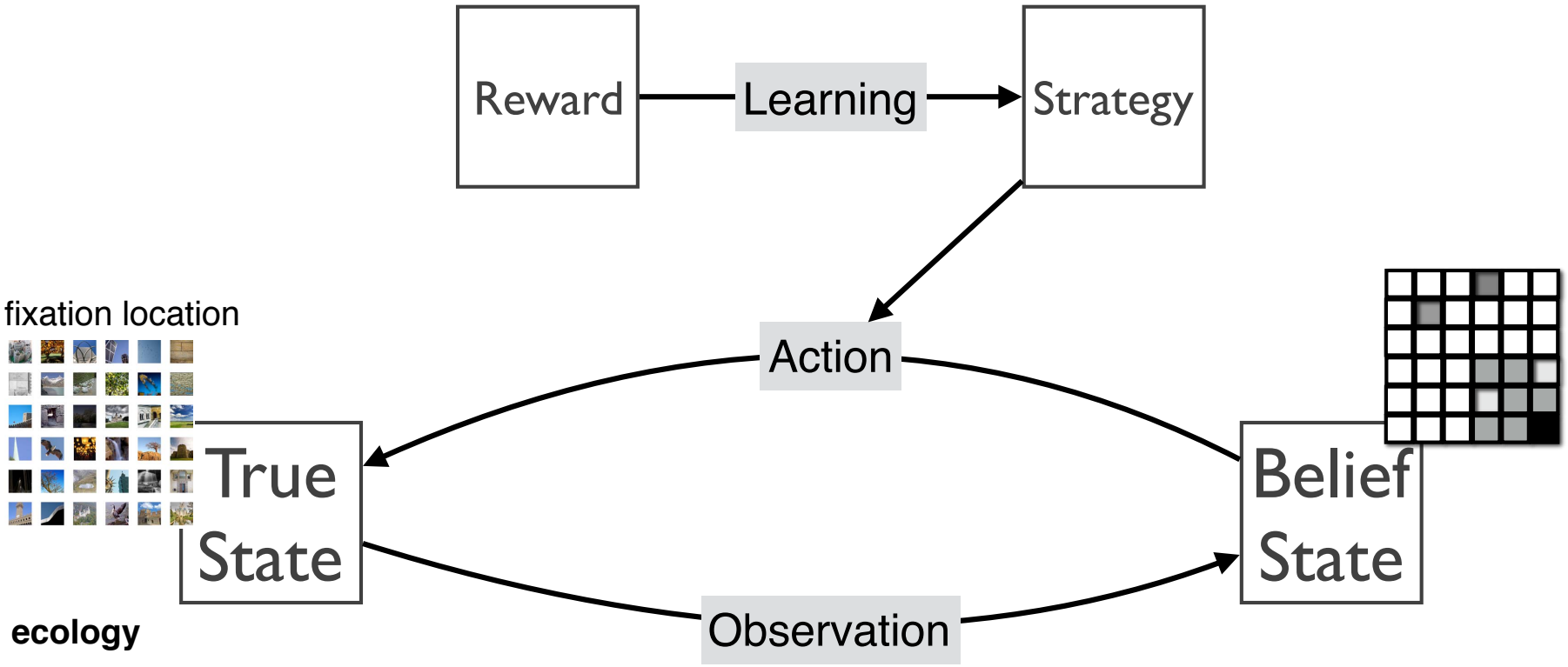
Andrew Howes
Summe School 2017

- a visual search task:
- find an image with the following features:
- sky, building, water, trees, reflection.



Tseng, Y. C., & Howes, A. (2015). The adaptation of visual search to utility, ecology and design. International Journal of Human-Computer Studies, 80, 45-55.

speed/accuracy trade-offs



ecology

The observation function

- The observation function is used to make a sequence of observations of the state.
- Observations are not of the whole state.
- Observations are not guaranteed to be correct.
- Observations result in a belief about the world that makes a commitment to one interpretation. The belief is a model.
- Varies across individuals.

The action function

- Important actions include: saccade and fixation.
- There are minimum time requirements for how long it takes to make eye movements
- For reading,
- Saccade duration is between150 and 175ms (Rayner, 2008).
- Fixation durations vary between 100 and over 500ms.
- where a millisecond (ms) is 1/1000th of a second.

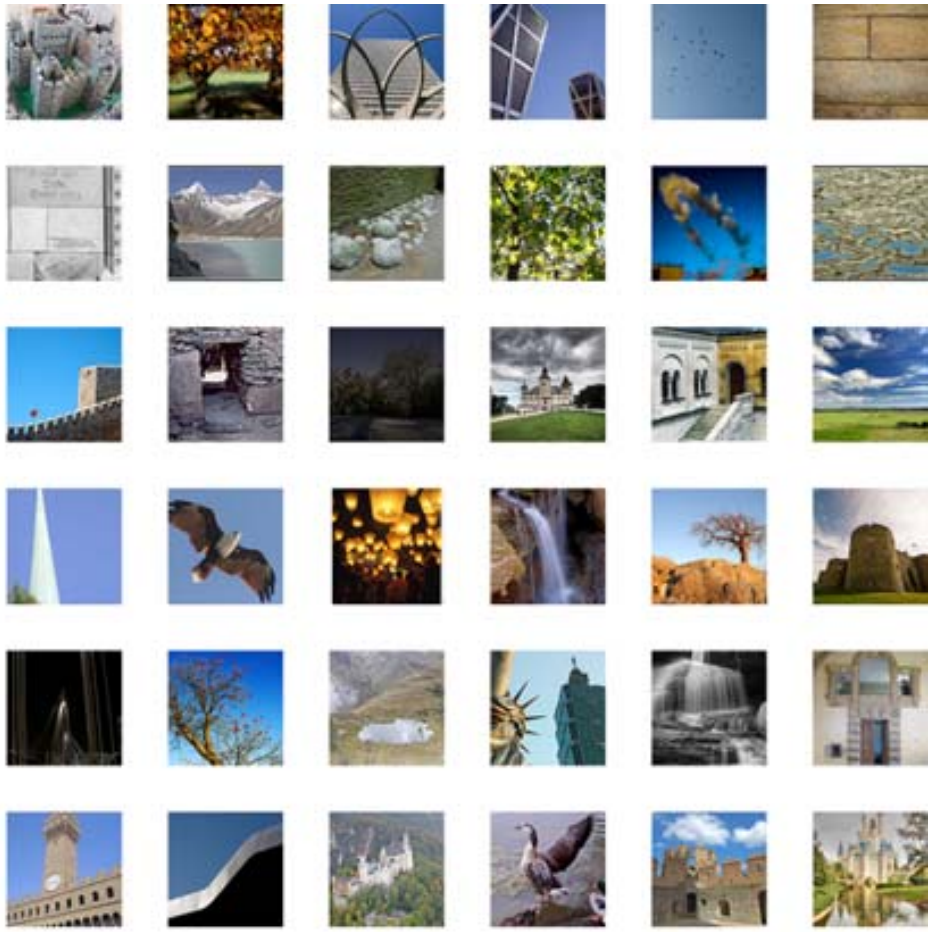
reward function

- The reward function represents the costs and benefits of action to the person.
- Information gain is one key benefit of fixations.
- All actions have a time cost.
- The benefit of saccades are deferred.
- People gain information from eye movements.

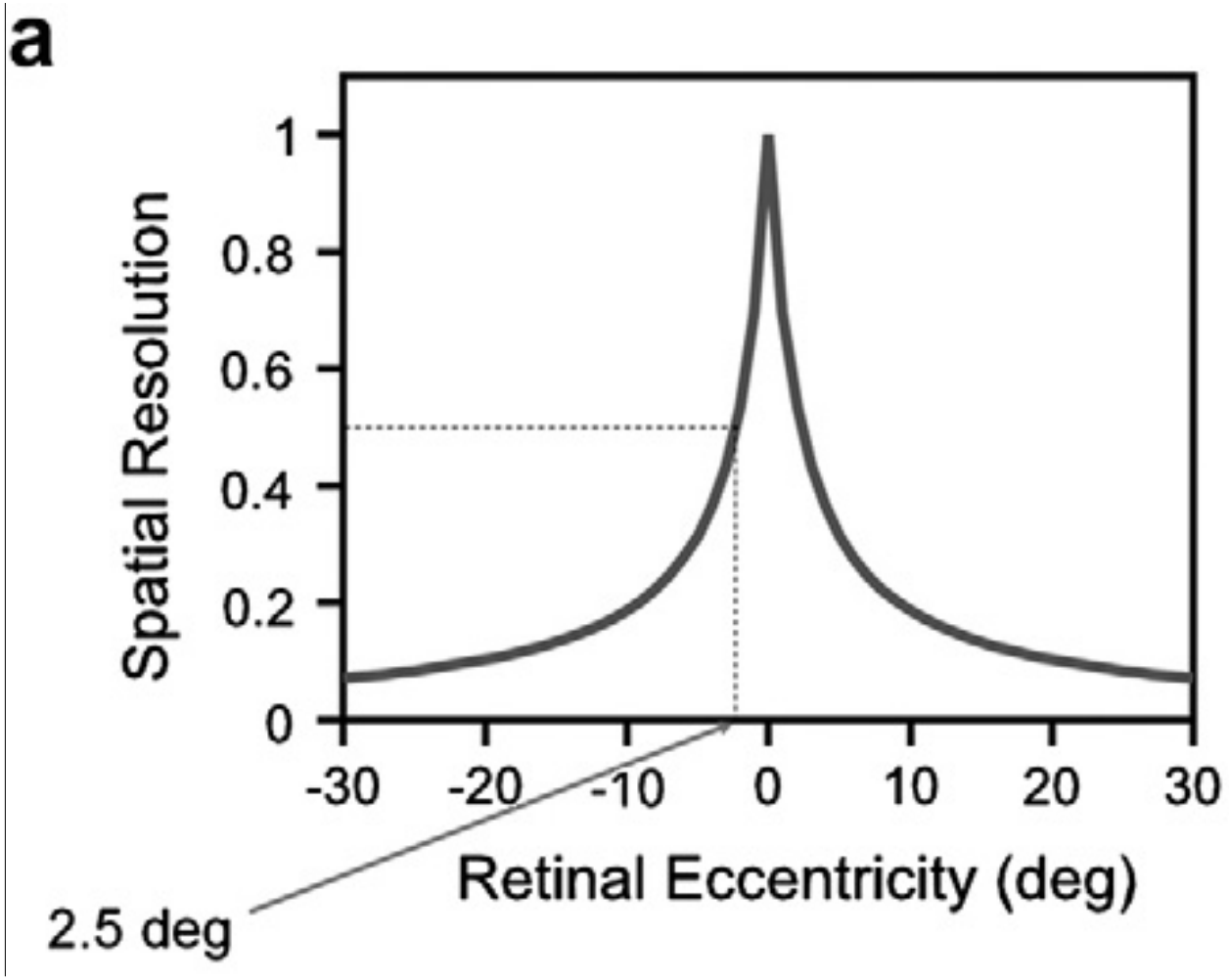
Multi-attribute reward functions

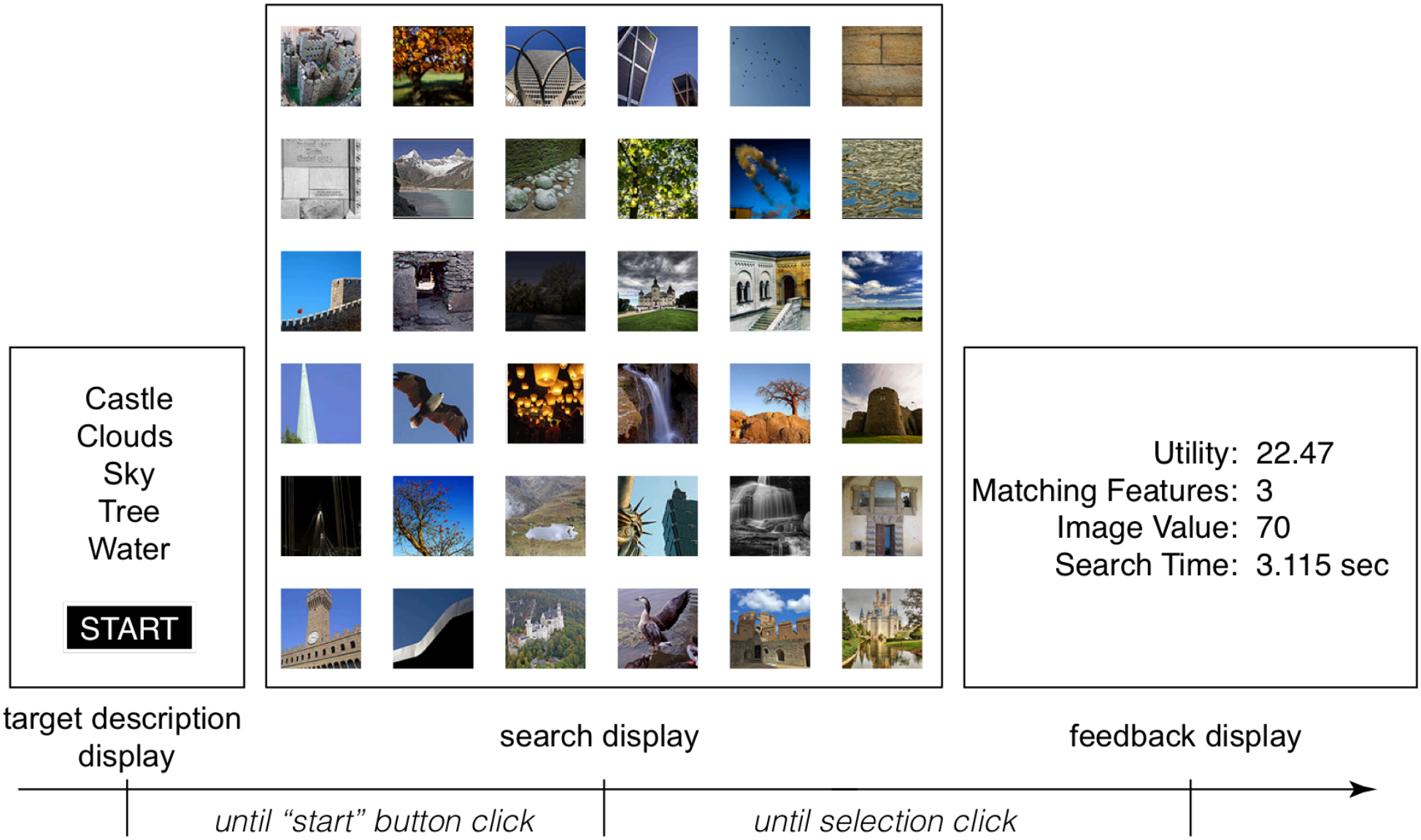
- Multi-attribute utility function for speed and accuracy
  - utility = speed x W1 + accuracy x W2
- Time and money.
  - utility = money x W1 - time x W2
- How people combine multiple attributes is an open research question (Vlaev, Seymor, Dolan, Chater, 2009; Talmi, 2009).

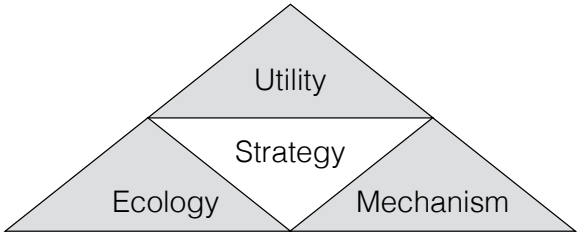
- What is the best layout of images?
- Larger or smaller?
- Spaced or adjacent?
- People adapt eye movements to the layout.

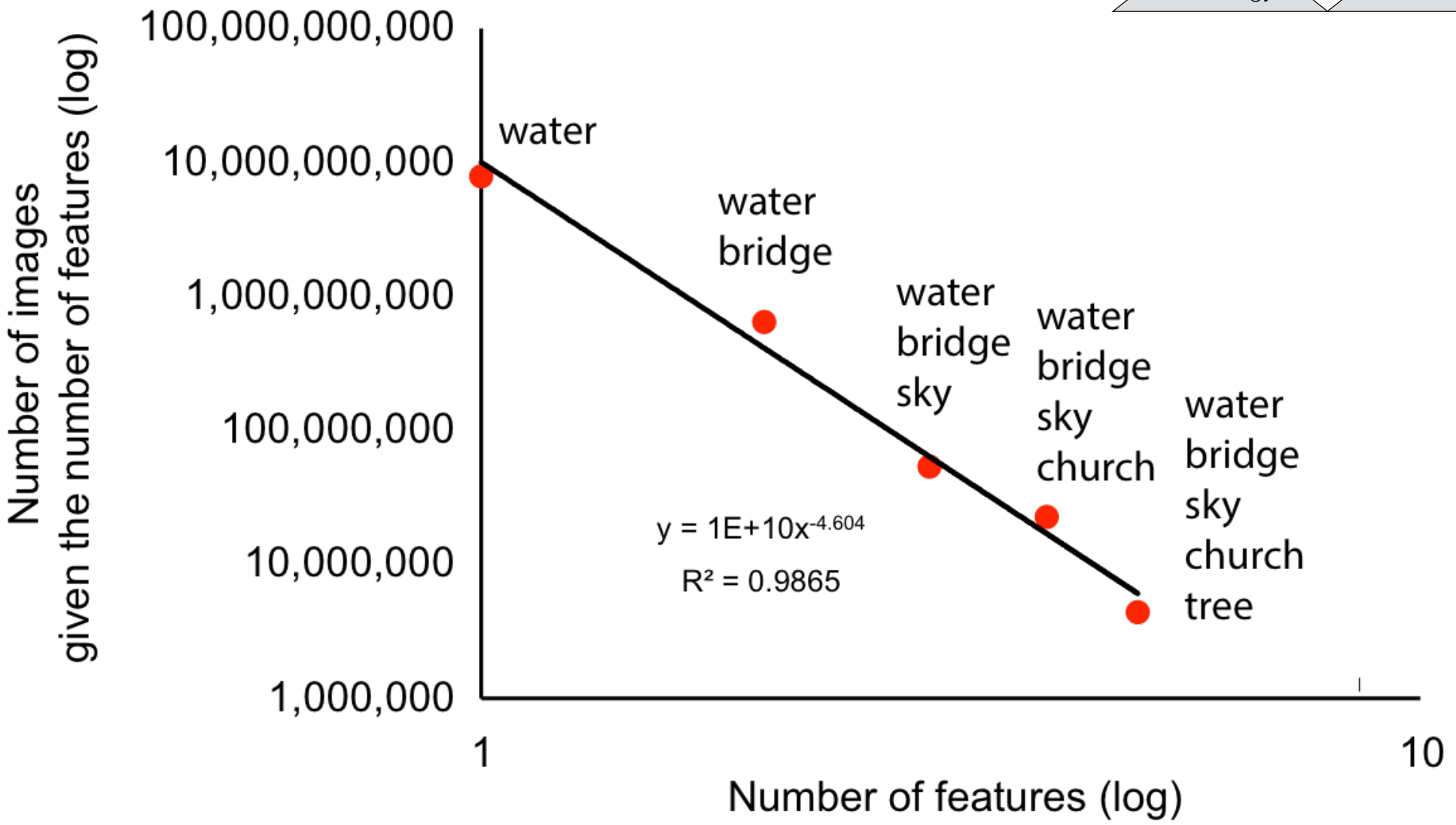


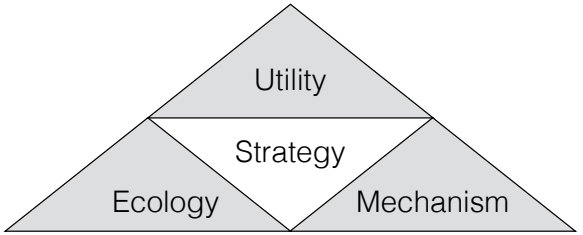
interference through crowding versus peripheral visual search

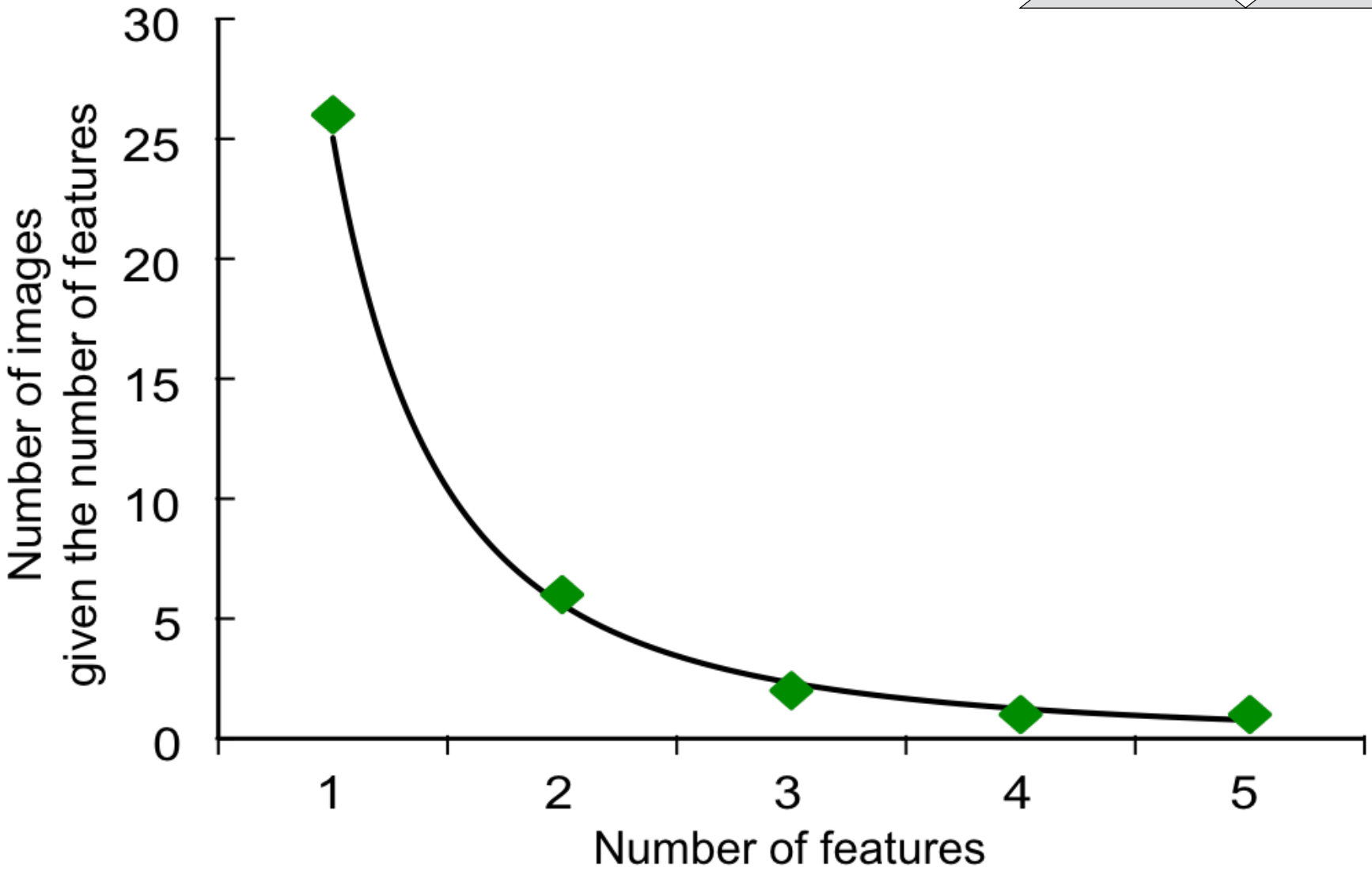


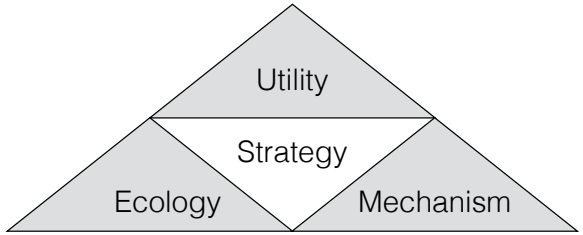

Castle
Clouds
Sky
Tree
Water
START
Utility: 22.47
Matching Features: 3
Image Value: 70
Search Time: 3.115 sec
target description display
search display
feedback display
until “start” button click
until selection click

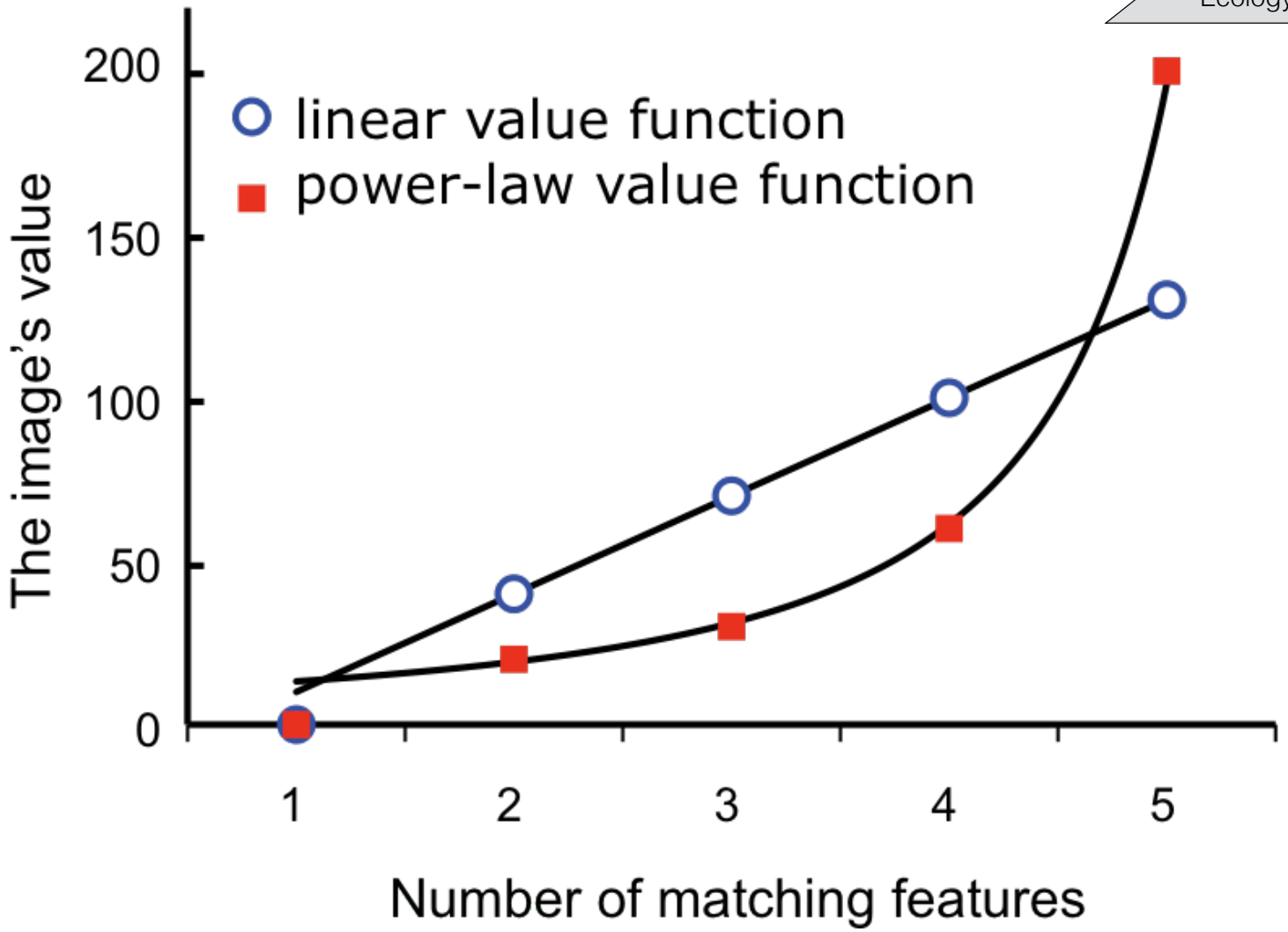

Utility
Strategy
Ecology
Mechanism


100,000,000,000
10,000,000,000
1,000,000,000
100,000,000
10,000,000
1,000,000
Number of images
given the number of features (log)
water
water
bridge
water
bridge
sky
water
bridge
sky
church
water
bridge
sky
church
tree
y = 1E+10x-4.604
R² = 0.9865
1
10
Number of features (log)

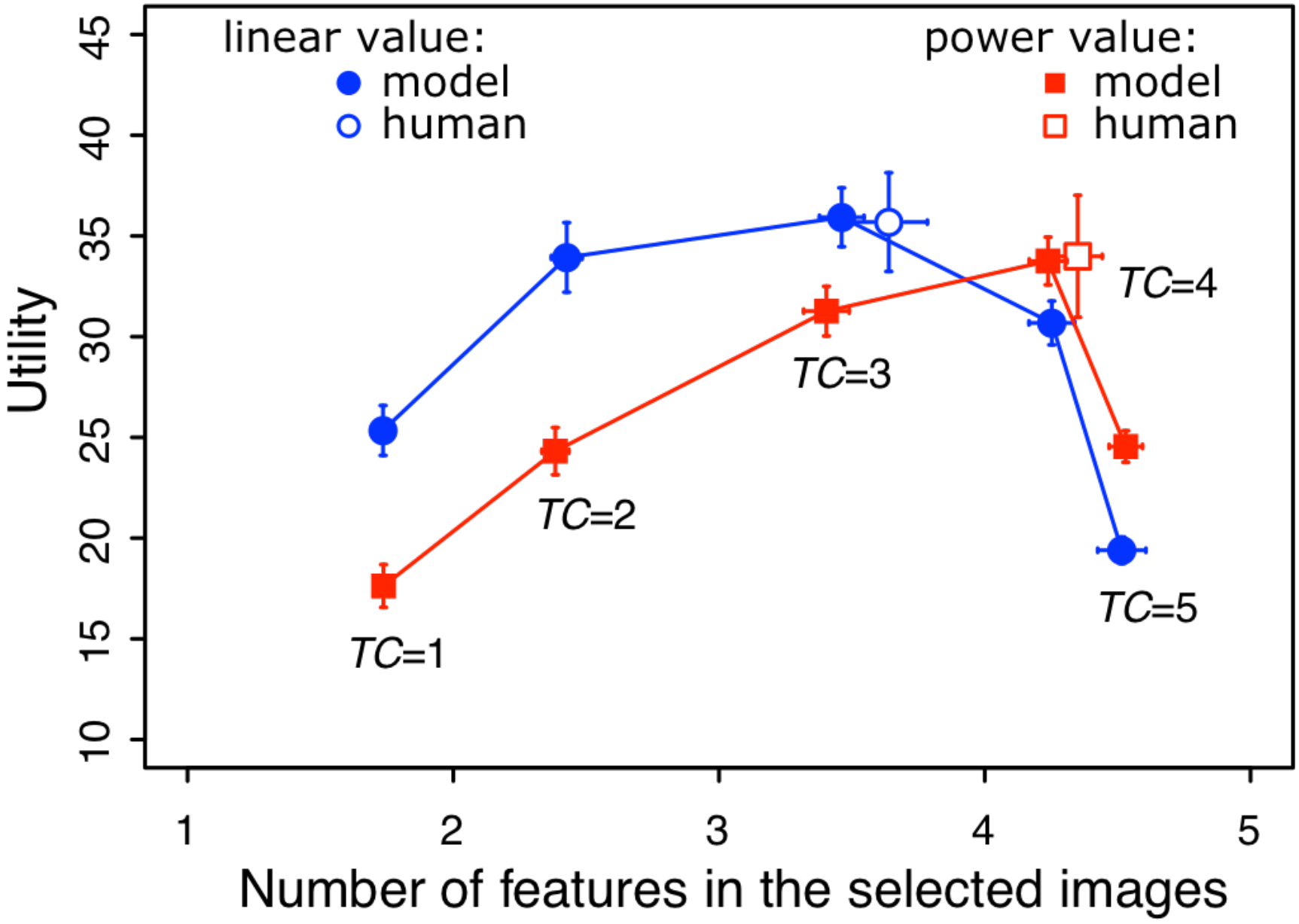

Utility
Strategy
Ecology
Mechanism

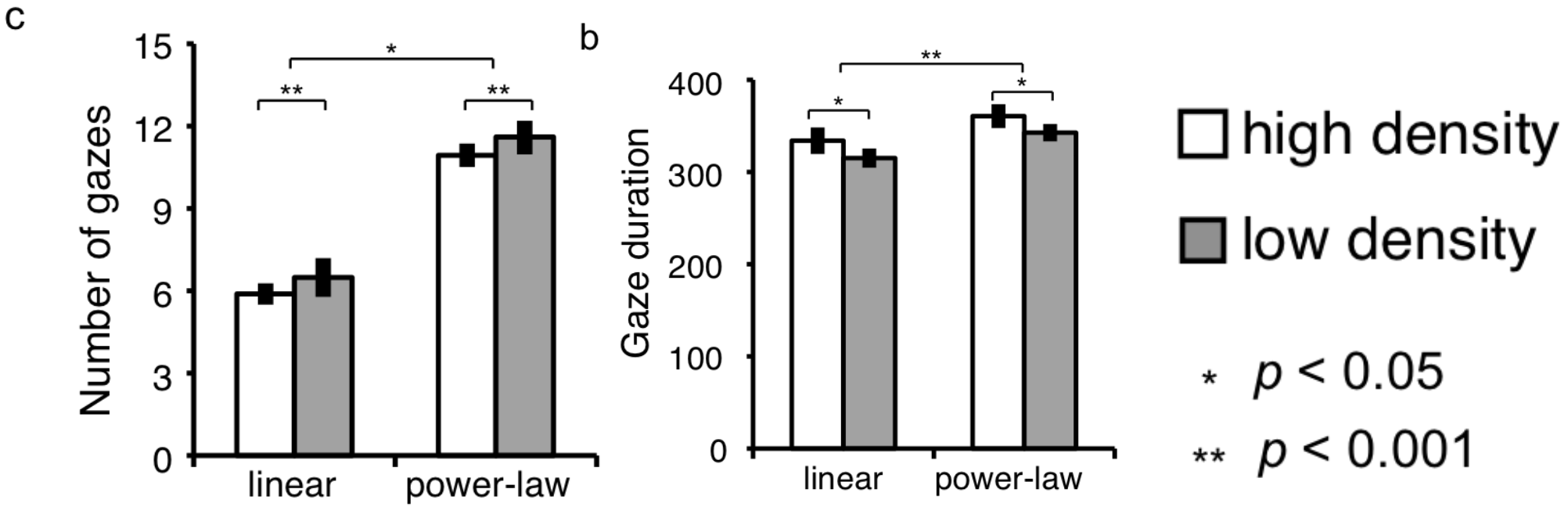

30
25
20
15
10
5
0
Number of images
given the number of features
1
2
3
4
5
Number of features


Utility
Strategy
Ecology
Mechanism


200
150
100
50
0
The image's value
linear value function
power-law value function
1
2
3
4
5
Number of matching features

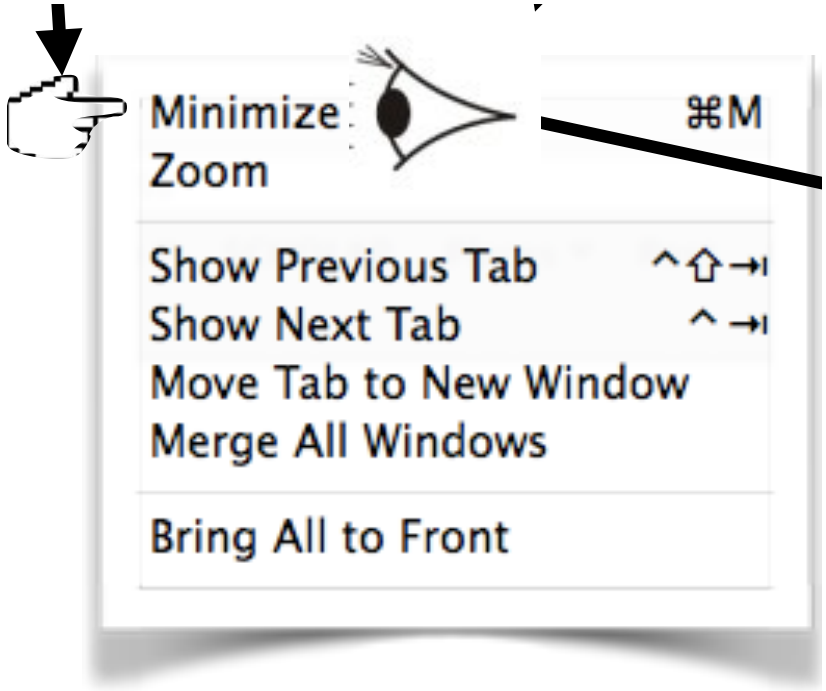
results


linear value:
model
human
power value:
model
human
Utility
10
15
20
25
30
35
40
45
1
2
3
4
5
Number of features in the selected images
TC=1
TC=2
TC=3
TC=4
TC=5

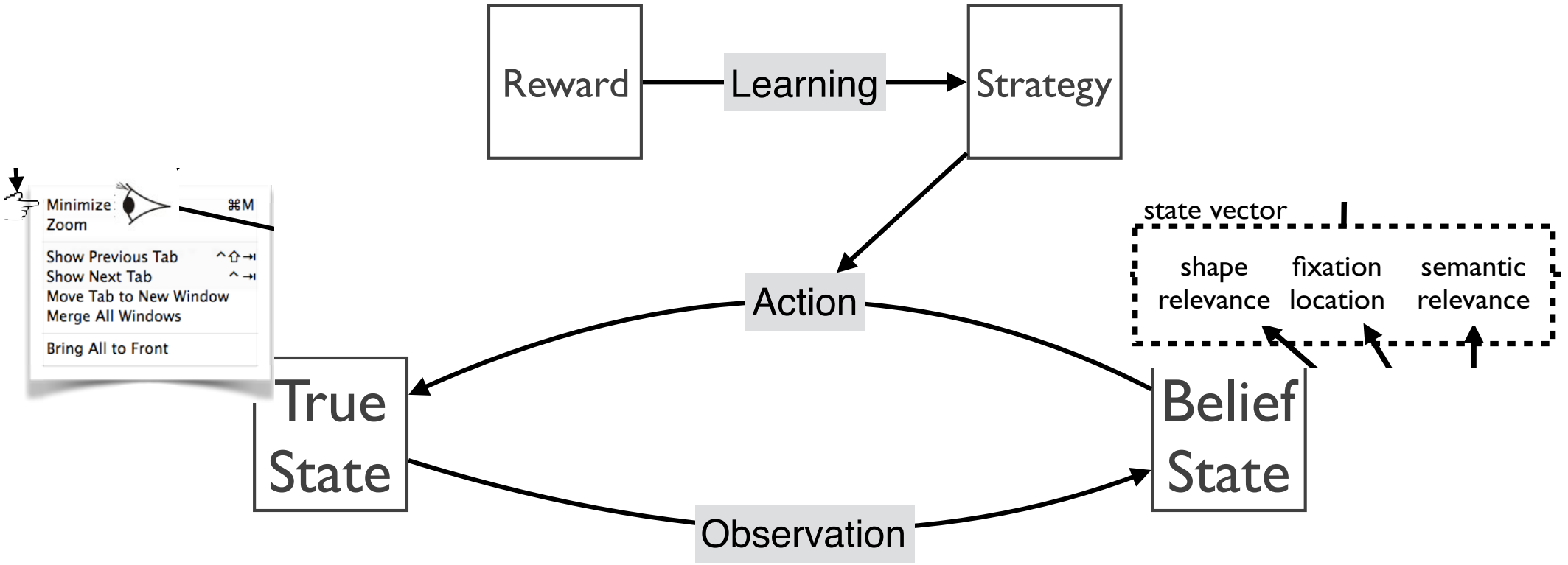

c
Number of gazes
15
12
9
6
3
0
*
**
**
linear
power-law
b
Gaze duration
400
300
200
100
0
**
*
*
linear
power-law
high density
low density
* p < 0.05
** p < 0.001

Chen, X., Bailly, G., Brumby, D. P., Oulasvirta, A., & Howes, A. (2015, April). The Emergence of Interactive Behavior: A Model of Rational Menu Search. *In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 4217-4226). ACM.

Menu search


Minimize ⌘M
Zoom
Show Previous Tab ^⇧⇥
Show Next Tab ^⇥
Move Tab to New Window
Merge All Windows
Bring All to Front

Menu Search as a POMDP



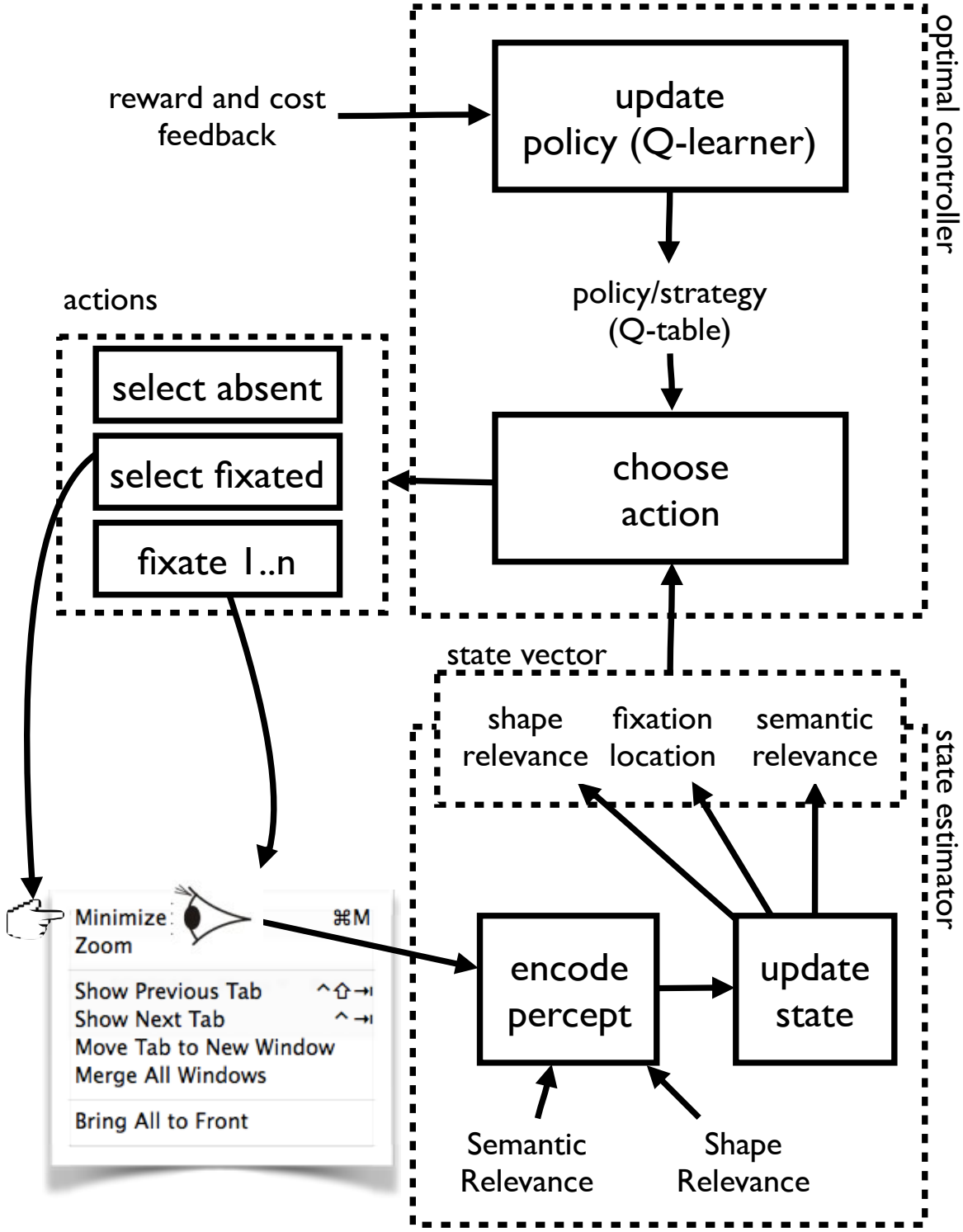
- Example changes in the state representation for a menu consisting of the first 4 items of the Safari Windows menu.

- State=[semantic relevance, fixation, shape relevance]

- Start with no knowledge about the menu(N = null).

- Given the goal of selecting "Show Next Tab" (item 4)

Action	*Semantic relevance*	*fixation*	*Shape relevance*
start	N, N, N, N	N	N, N, N, N
fixate 1	$0, N, N, N$	1	$0, 0, N, N$
fixate 3	$0, N, 0.3, N$	3	$0, 0, 1.0, 1.0$
fixate 4	$0, N, 0.3, 1.0$	4	$0, 0, 1.0, 1.0$
select 4			


reward and cost feedback
update policy (Q-learner)
optimal controller
policy/strategy (Q-table)
actions
select absent
select fixated
fixate 1..n
choose action
state vector
shape relevance
fixation location
semantic relevance
state estimator
Minimize ⌘M
Zoom
Show Previous Tab
Show Next Tab
Move Tab to New Window
Merge All Windows
Bring All to Front
encode percept
update state
Semantic Relevance
Shape Relevance

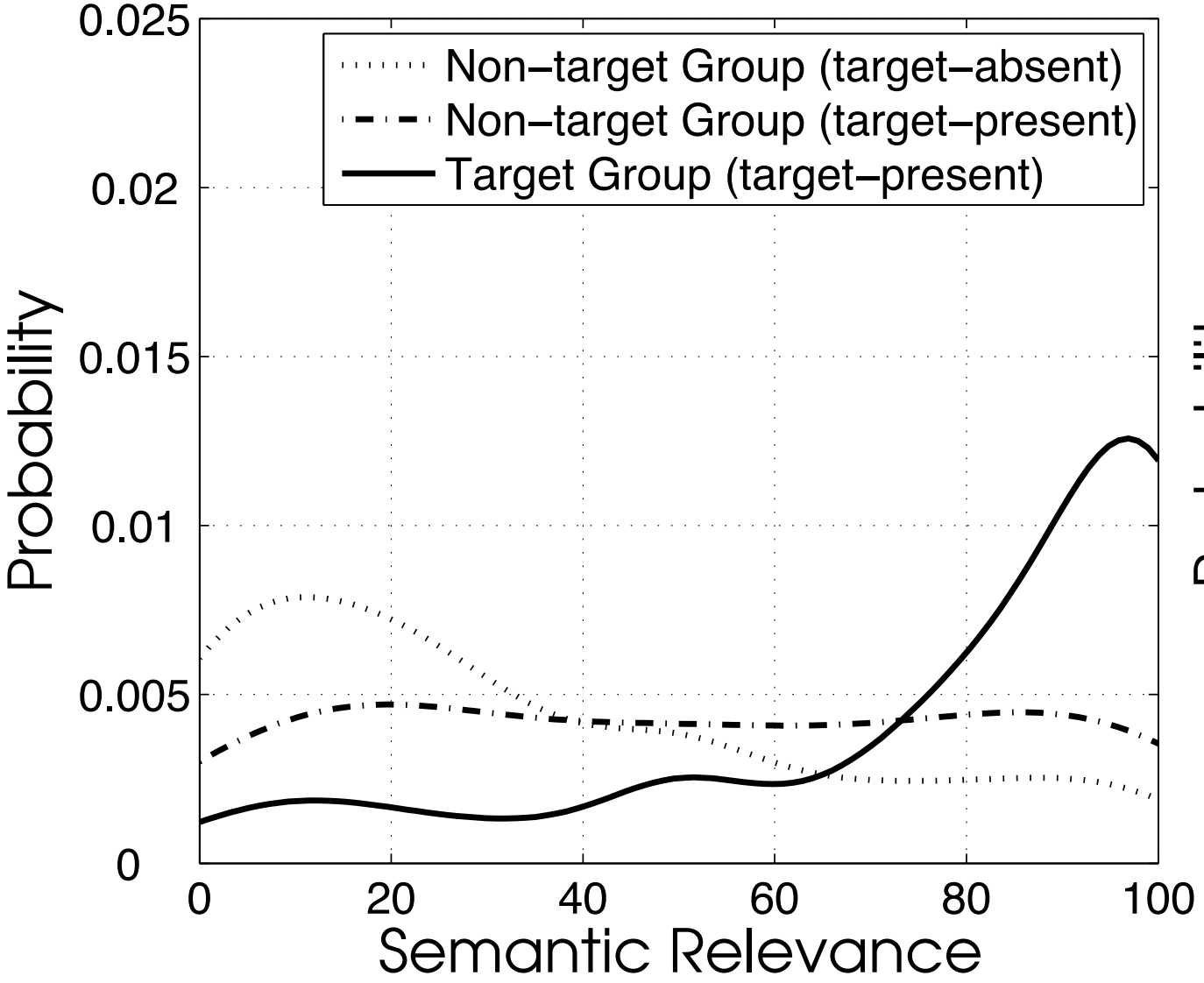
Evaluation

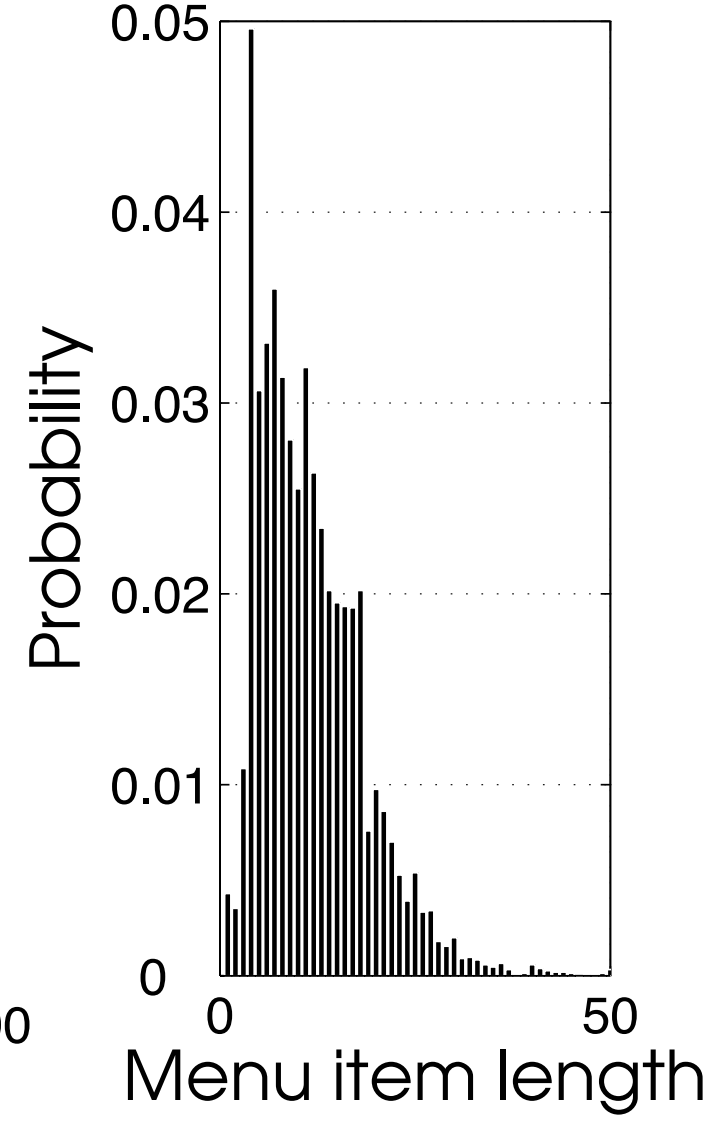
- Study 1: A sample of 600 real world Apple Mac application menus
- Study 2: Menus from an experiment with humans (Bailly et al. 2014)

Study 1: Model Implementation

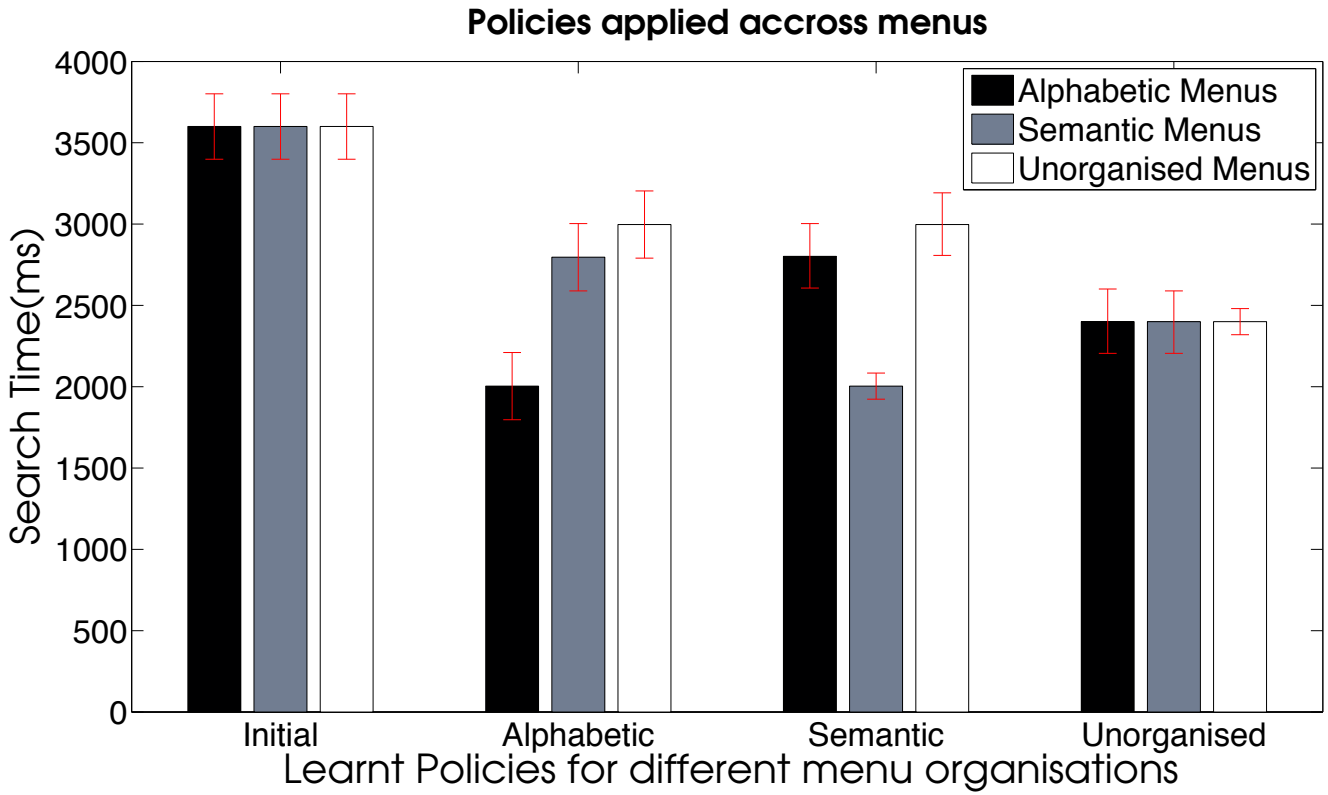
- Training phase was given 20 million trials to acquire the optimal policy.
- Test phase: the optimal policy was run on a further 10,000 trials of newly sampled menus, and its performance was recorded.
- In the Results, we extracted various aggregate performance measures.

Study 1: Apple Menus

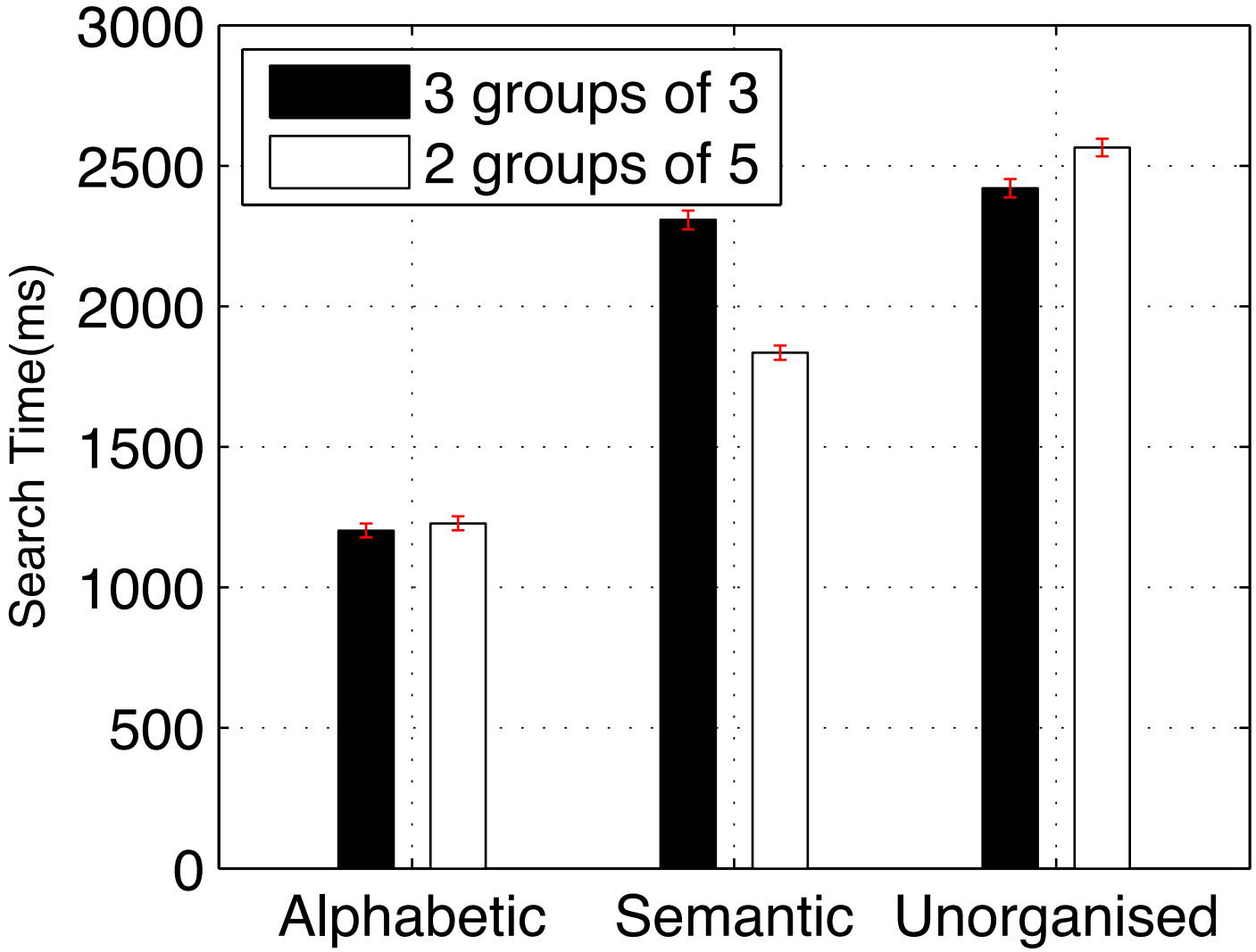




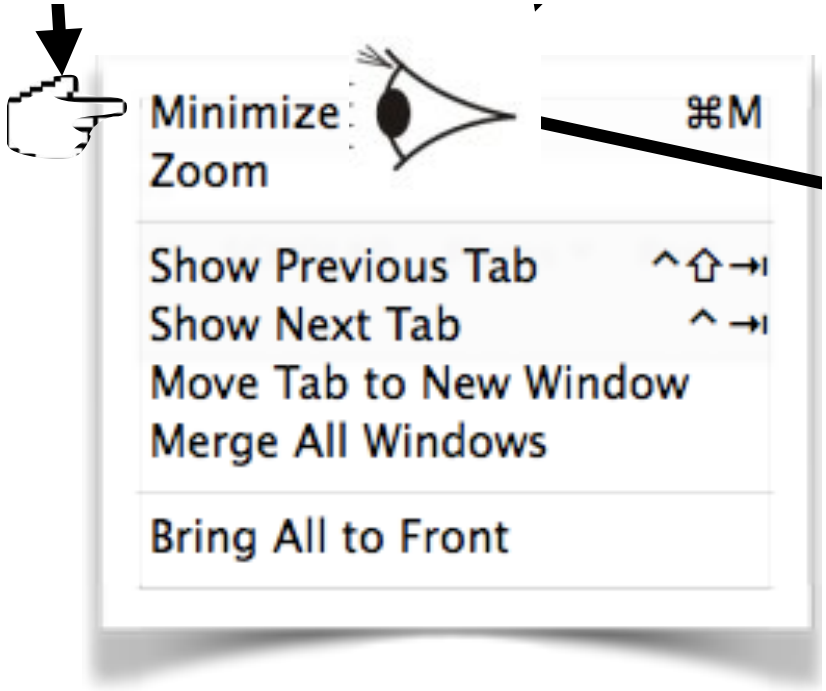
Study 1: search duration



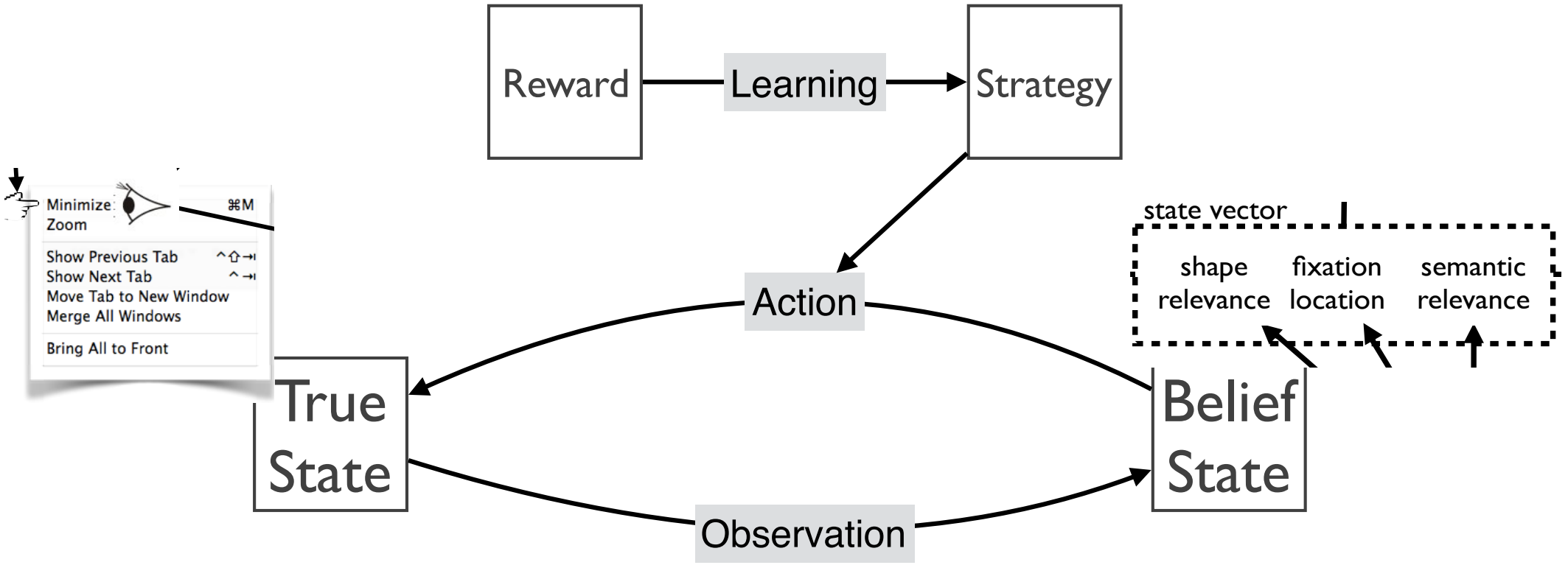
Study 1: Effect of semantic groups



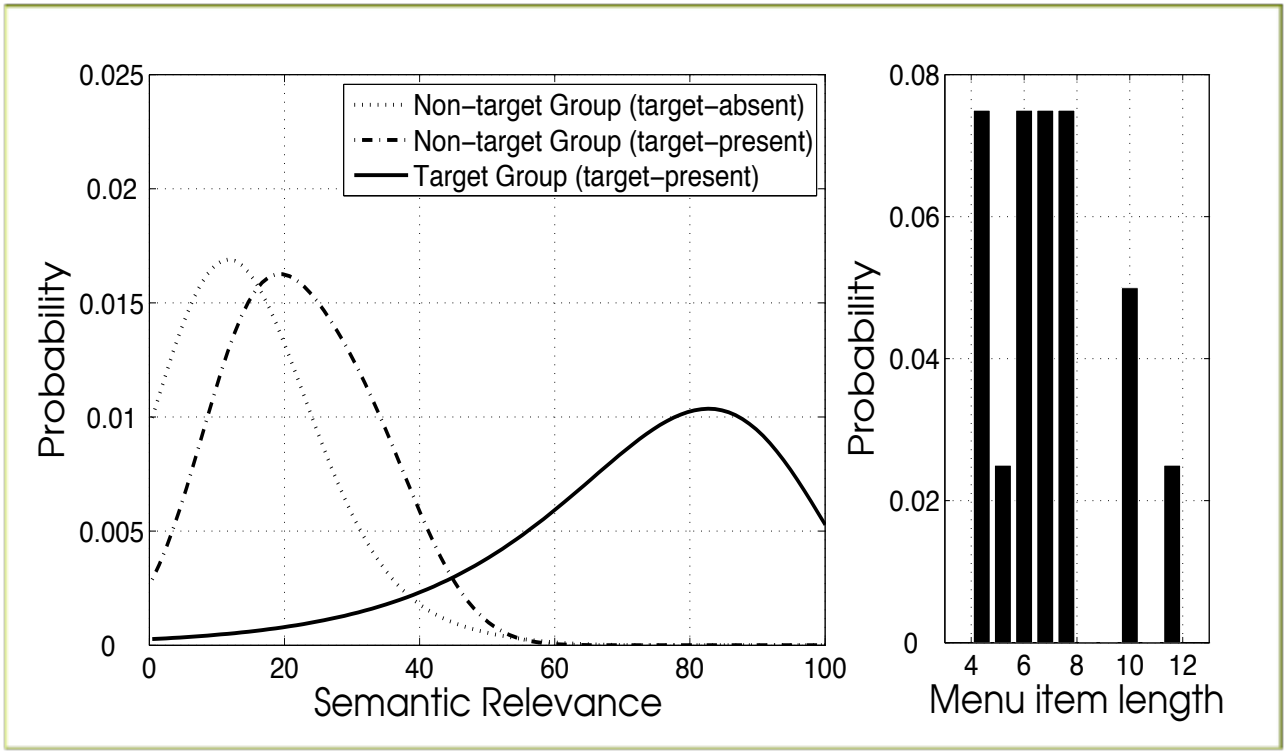
Menu search


Minimize ⌘M
Zoom
Show Previous Tab ^⇧⇥
Show Next Tab ^⇥
Move Tab to New Window
Merge All Windows
Bring All to Front

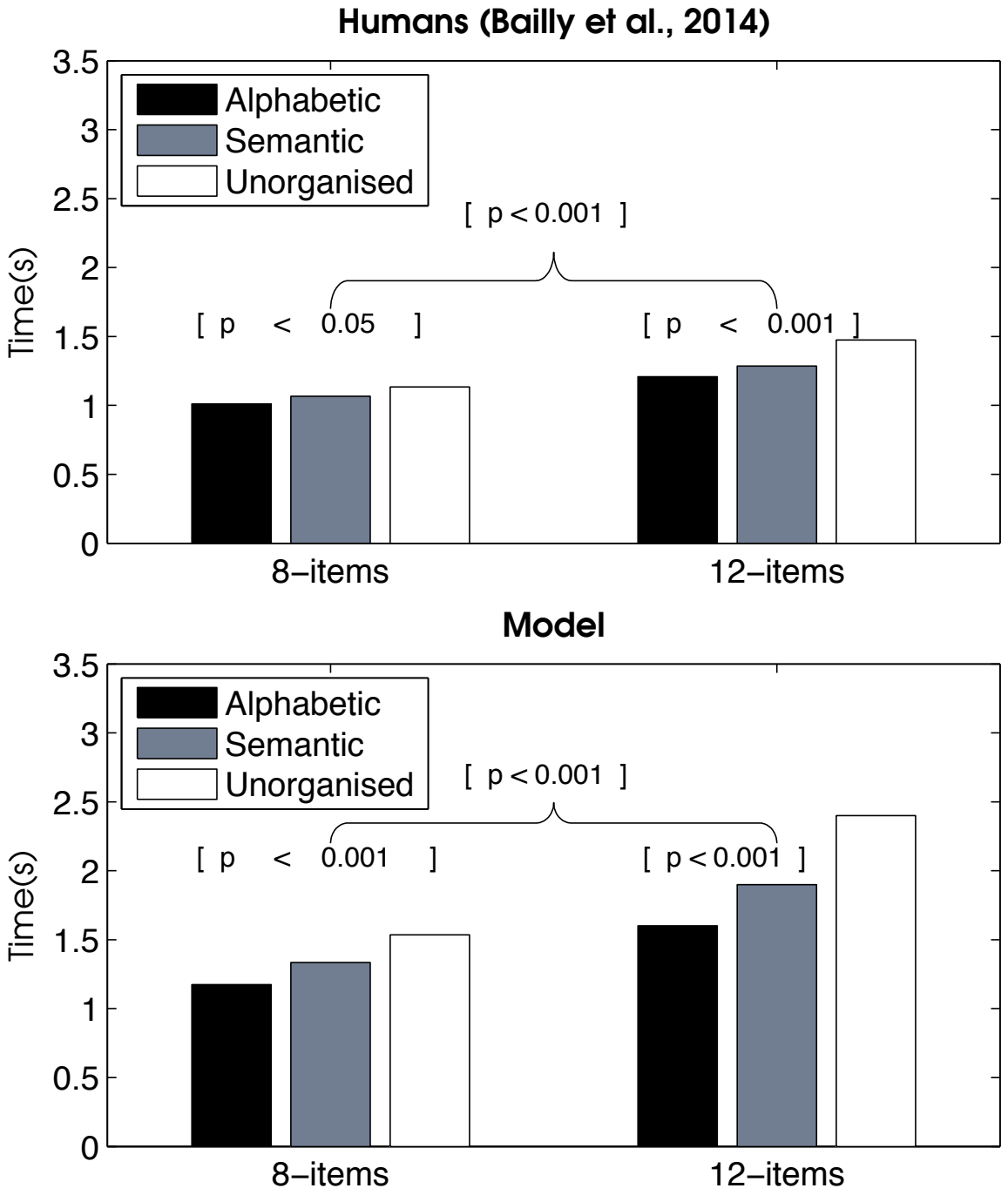
Menu Search as a POMDP



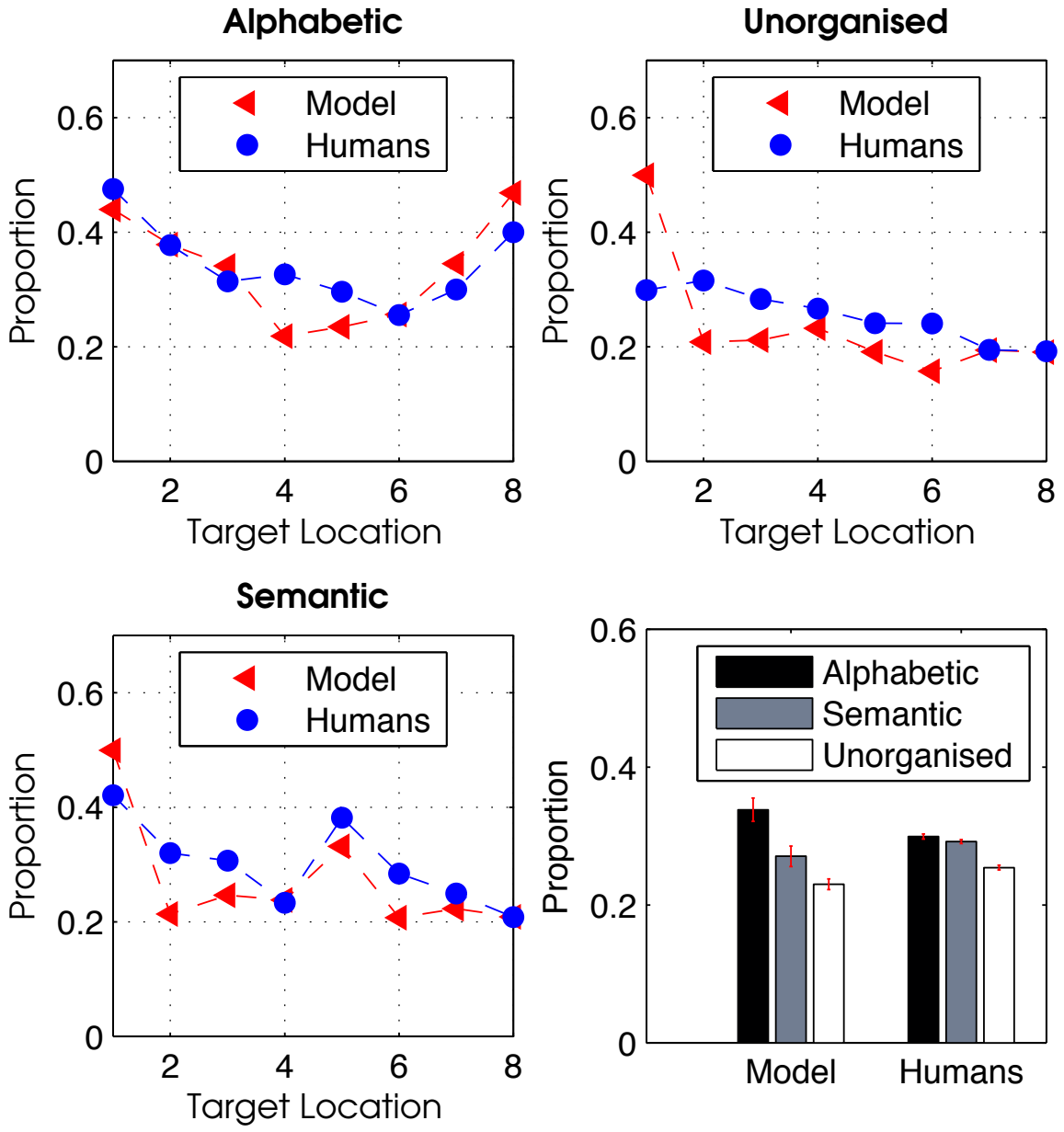
Study 2: Bailly et al. (2014)



Study 2: Search duration



Gaze allocation



Discussion

- We have demonstrated that interactive behaviour can be predicted as an emergent consequence of underlying constraints.
- A model of menu search predicted phenomena including: search time, gaze allocation, effect of organsation etc..
- The approach offers a radical alternative to a number of other approaches to cognitive modeling in HCI that are achieved by virtue of the fact that the predictions are an emergent consequence of defining the user's optimal control problem.

References

- Rayner, K (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, vol. 124, pp. 372–422.
- Kieras, D. E., & Hornof, A. J. (2014). Towards accurate and practical predictive models of active-vision-based visual search. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3875-3884). ACM.
- Butko, N. J., & Movellan, J. R. (2009). Optimal scanning for faster object detection. In Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on (pp. 2751-2758). IEEE.