a Experiment and design

Task

Procedure

Tutorial round

8 rounds of search

25 clicks in each round

"Bonus round" with reward predictions
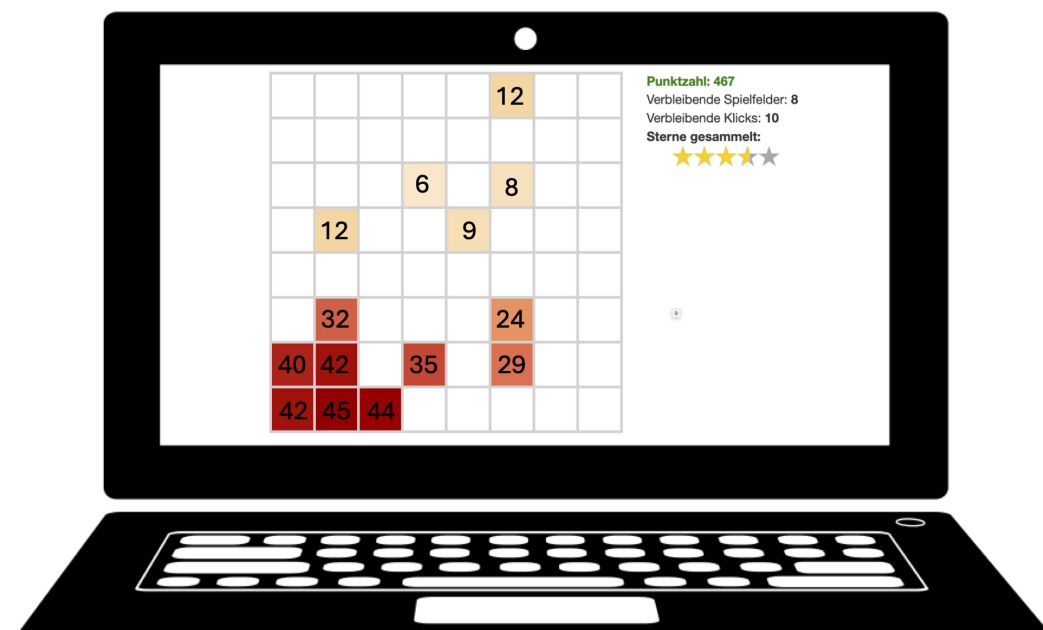
Time
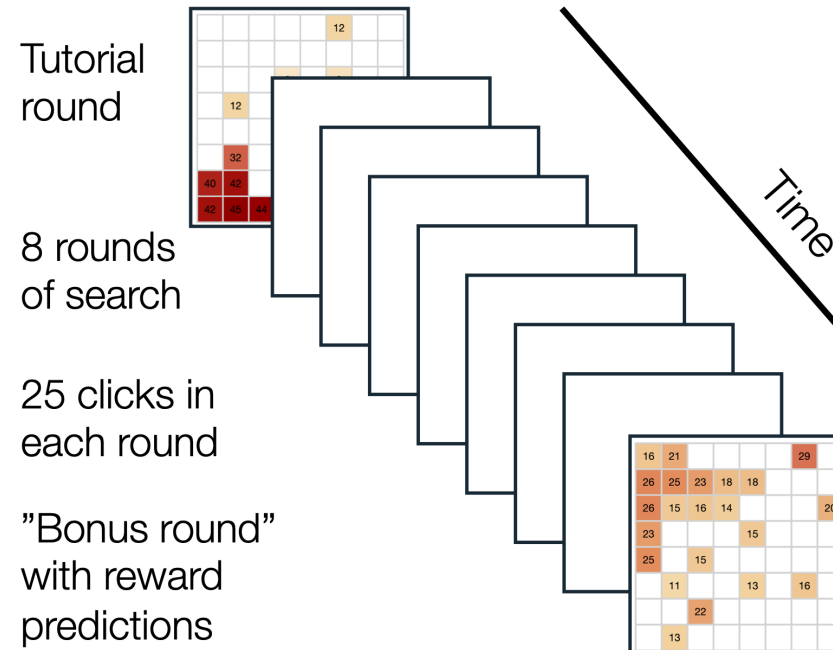
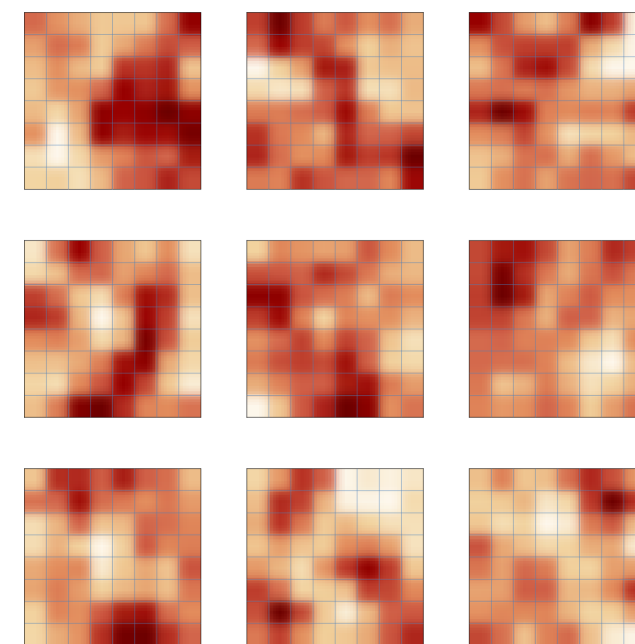## Environments

b Computational GP–UCB model

Observations (t=8)
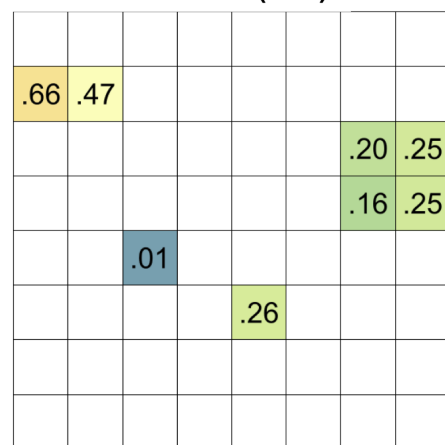
Expected rewards
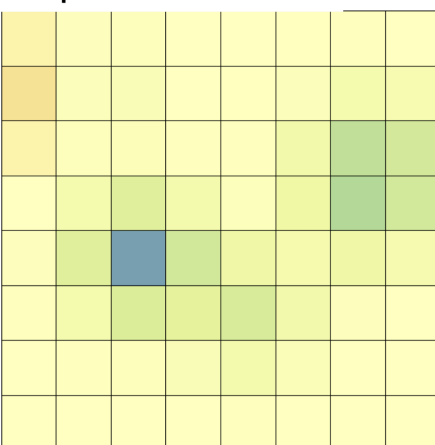
Uncertainty
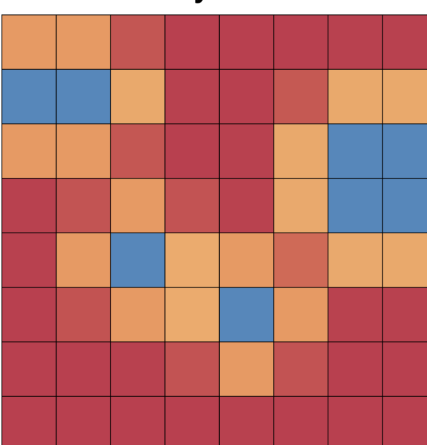
Upper Confidence Bound (UCB)
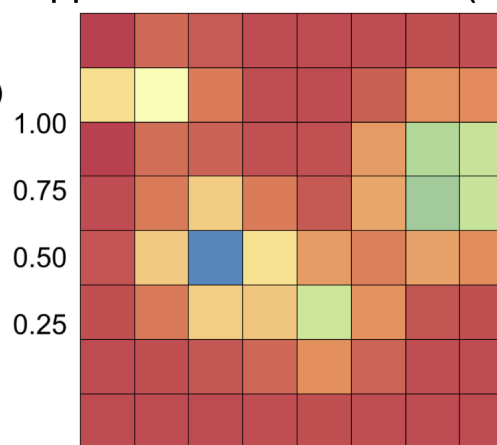
Softmax predictions

$\mu(\mathbf{x})$

$\sigma(\mathbf{x})$

$UCB(\mathbf{x})$

$p(\mathbf{x})$

Generalization λ

Exploration bonus β

Random temperature τ