Evaluating ideal observers for large target identification tasks under additive white noise

Can Oluk[1], Wilson S. Geisler[2]

1 Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
2 Center for Perceptual Systems and Department of Psychology, University of Texas at Austin

Bayesian Ideal Observers

performs a given task at the optimal level possible, given the available information and any specified constraints (Geisler, 2011).

Advantages:

- Almost parameter-free, quantitative predictions
- Have provided important insights into visual processing
- Provide principled benchmark predictions
- Specify optimal computations and task-relevant information
- Guide principled modeling of suboptimal performance

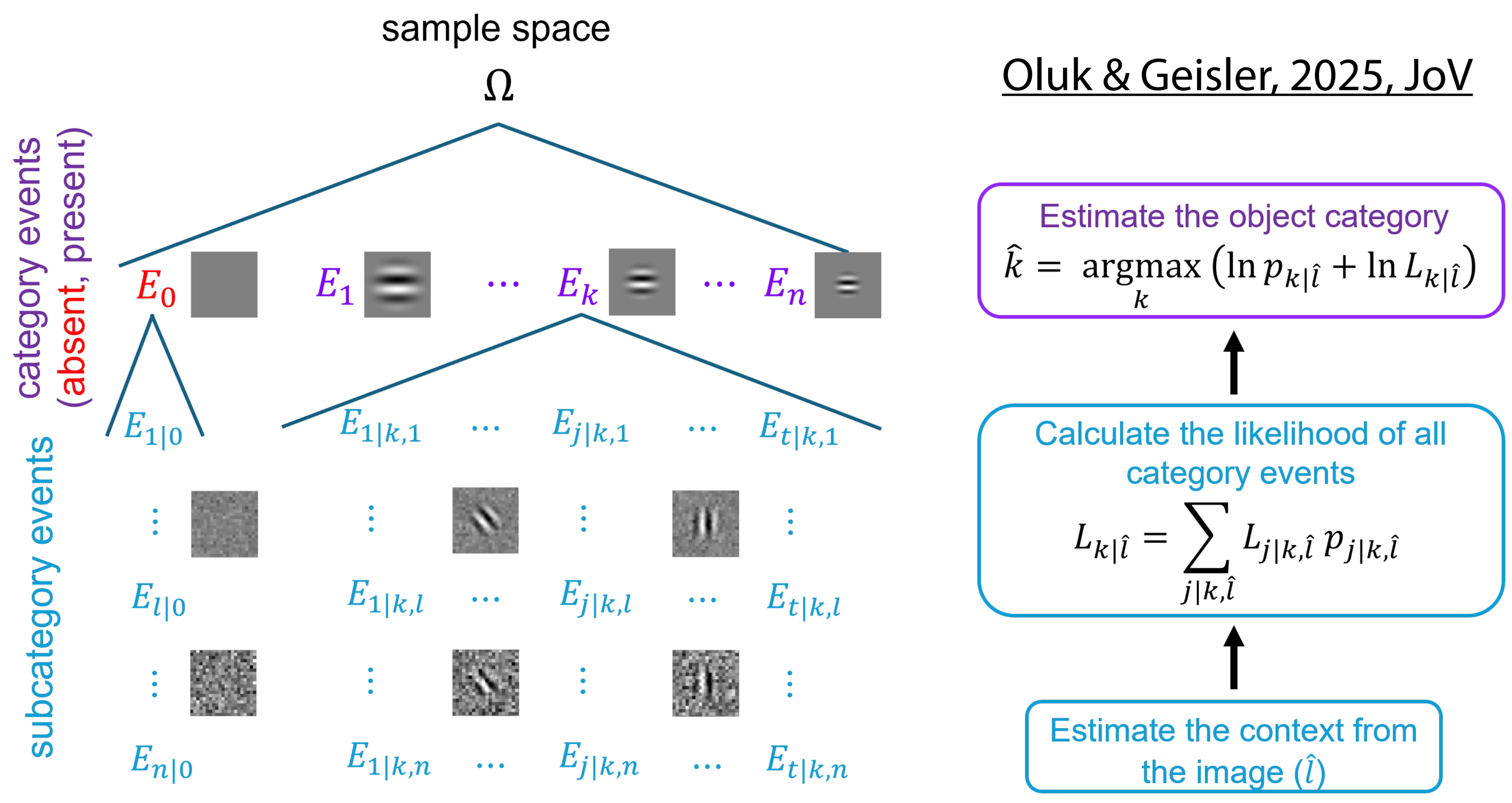
However, evaluating ideal observers for large, complex tasks is challenging.

We aim to scale up the set of tasks for which ideal observers can be evaluated.

Target Identification

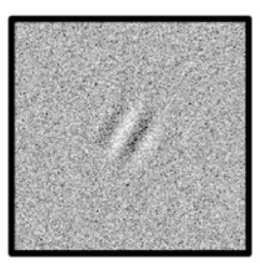
For **arbitrarily large and complex** target identification tasks:

- Describe the task as a hierarchy of exhaustive, mutually exclusive events.
- Derive general equations for ideal observers.



Toolbox

A toolbox for evaluating ideal observers under additive white noise: **github.com/CanOluk/Target_Identification**



$$\hat{k} = \arg\max_k \left(\ln p_{k|\hat{l}} + \ln \sum_{j|k,\hat{l}} p_{j|k,\hat{l}} \exp\left(d'_{j|k,\hat{l}} R'_{j|k,\hat{l}} - 0.5 d'^2_{j|k,\hat{l}} \right) \right)$$

1. Evaluate large tasks

A convolutional neural network does not achieve ideal performance, even with extensive training

2. Evaluate multiple tasks

(i) Signal-to-noise ratio (ii) Uncertainty levels (iii) Discrimination levels

(i) Prior probabilities (ii) Cost functions

Datasets

Sine waves

36 orientations x 11 scales
Total targets: 396
Total images: 200,000

Precomputation time: 2 min / 400 MB

Detection of target:
11 background contrast
x 16 target amplitudes (69,696)

Evaluation of the ideal observer: 6 s

CIFAR-100

20 superclass,10 subclass, 100 images
Total targets: 10,000
Total images: 190,000

Precomputation time: 8 min/ 13.6 GB

Detection of target:
3 background contrast
x 3 target amplitudes (90,000)

Evaluation of the ideal observer: 7 min

Conclusion 1

We leverage all advantages of evaluating ideal observers for large tasks.

Potential applications: (i) handwritten digits;
(ii) optical distortions (e.g., blur, aberrations) that induce internal variability.

Potential Extensions: additive filtered-noise backgrounds.

Basis Simulation

$\mathbf{R}' = [R'_{j|k,\hat{l}}]$ Vector of normalized-template responses for a single trial

$\mathbf{R}'(\mathbf{J}, \mathbf{K}, \mathbf{L})$ Vector when the image presented subcategory (J), category (K), and context level (L)

$\mathbf{B}'$ Vector when image without a target and standard deviation, nominally 1.0

$\mathbf{R}'(\mathbf{J}, \mathbf{K}, \mathbf{L}) = \mathbf{B}' + d_{J|KL'} \left[\mathbf{t}_{J|KL'} \cdot \mathbf{T}\right]$ Basis Equation

Pre-compute: $\mathbf{B}'$ large set of randomly sampled backgrounds; $\mathbf{t}_{j|k} \cdot \mathbf{t}_{i|k}$ all possible pairings of template responses

vs Alternatives

1. Monte Carlo Simulations

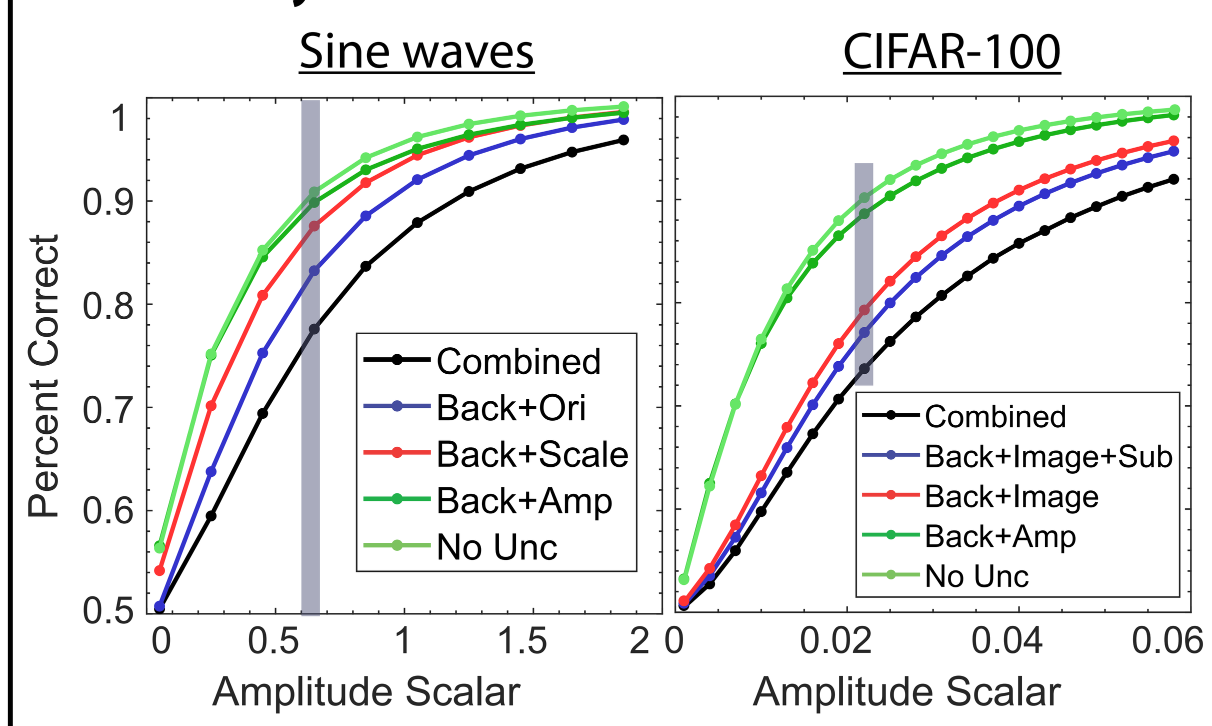
+ Avoids computing dot products between images and templates
- Requires convolving templates to compute the covariance matrix
- Requires recomputing the covariance for each condition to sample even a single trial

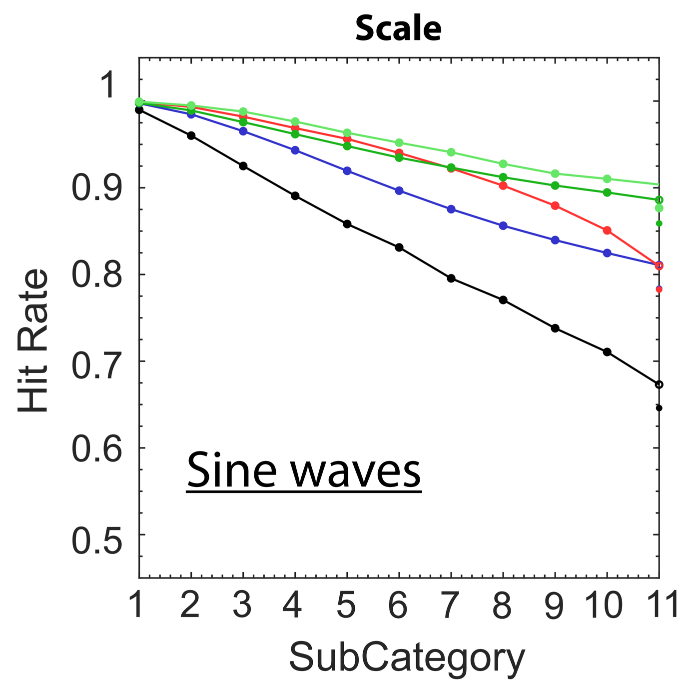
2. Analytical Approximations

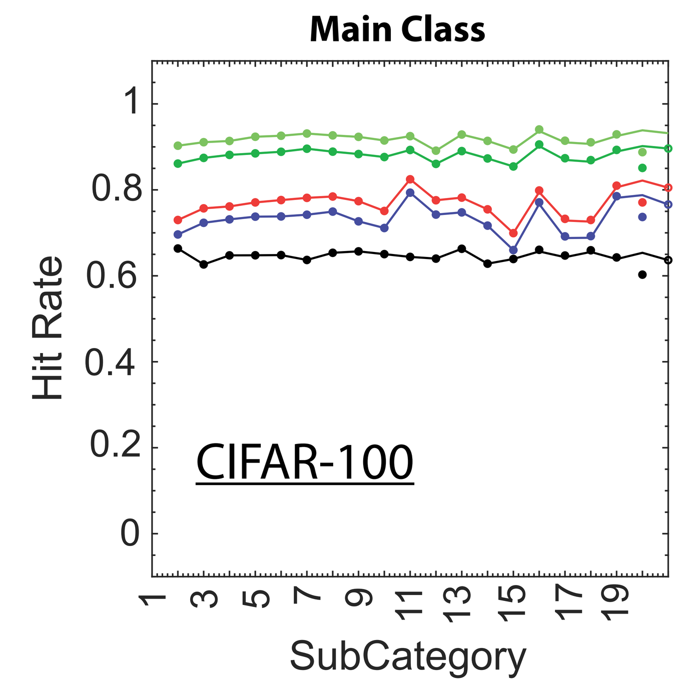
- Work only in limited cases; designed to approximate parts of the distribution

Results

Uncertainty levels

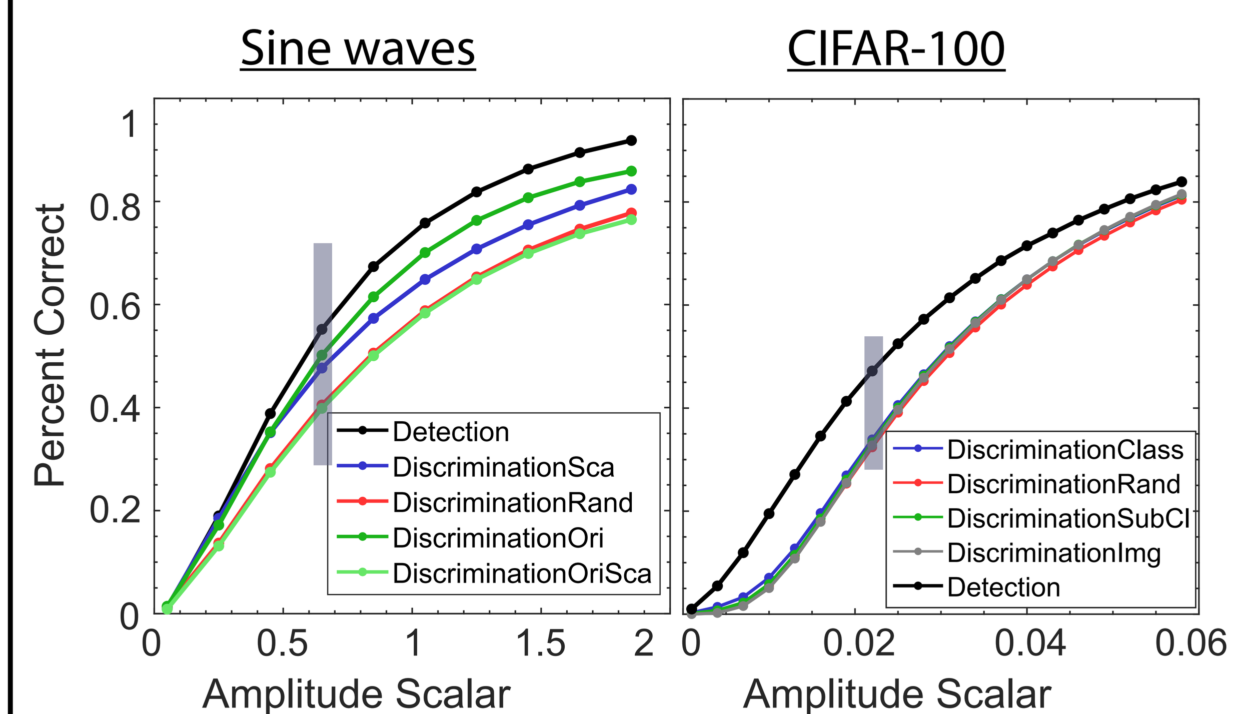


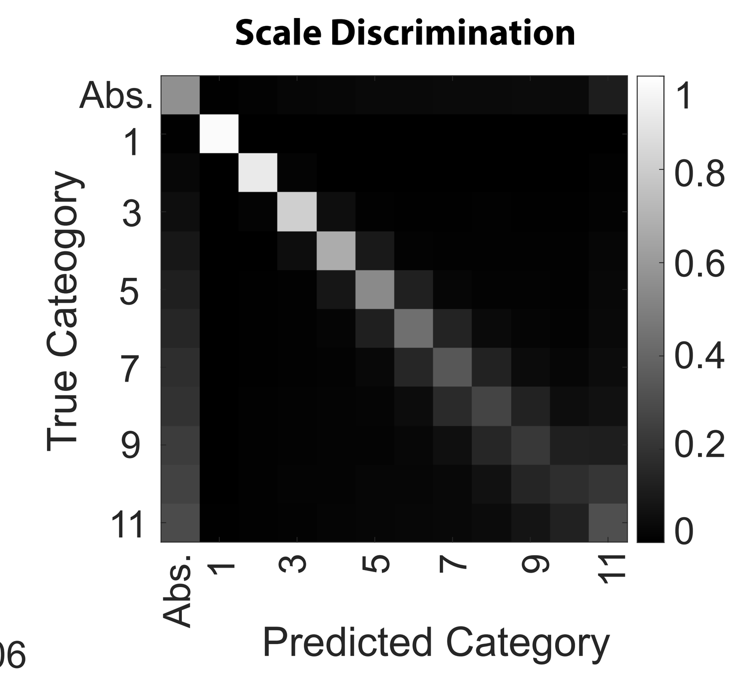


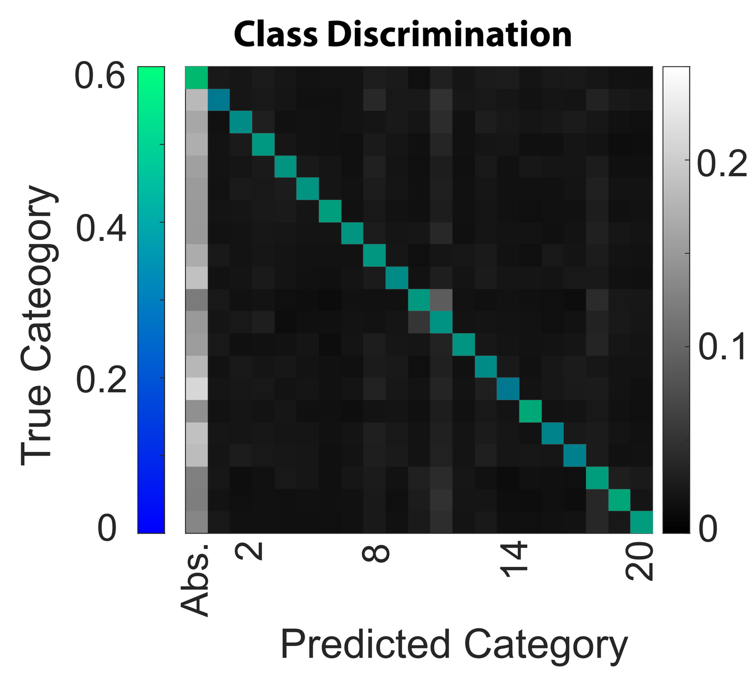


- Performance decreases as uncertainty increases.
- Rich predictions for hit and correct-rejection rates.

Discrimination levels







- Discrimination is harder than detection.
- Effects of specific dimensions depend on the dataset.
- Rich predictions for confusion matrices.

Conclusion 2

Our toolbox is particularly well suited to evaluating multiple tasks within a given dataset.

- Provides a deeper understanding of task constraints.
- Generates further testable predictions about relationships between tasks.

Prior change (Oluk & Geisler, 2023) Uncertainty levels (Oluk & Geisler, 2025) Discrimination levels (Bias: Poster at SVSS)