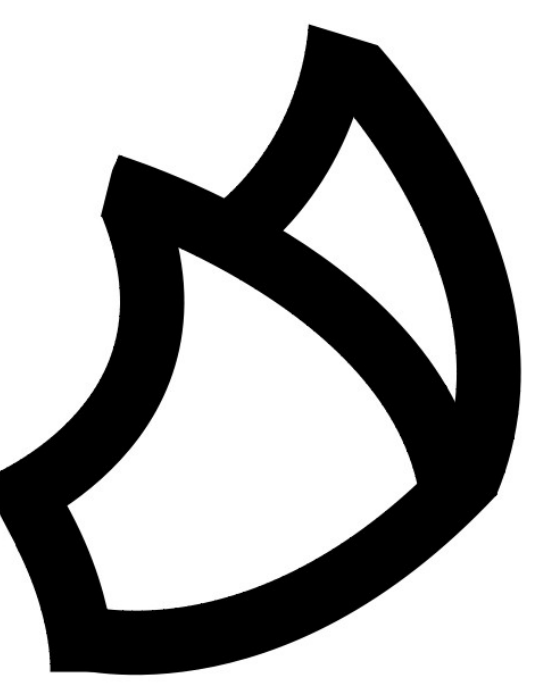




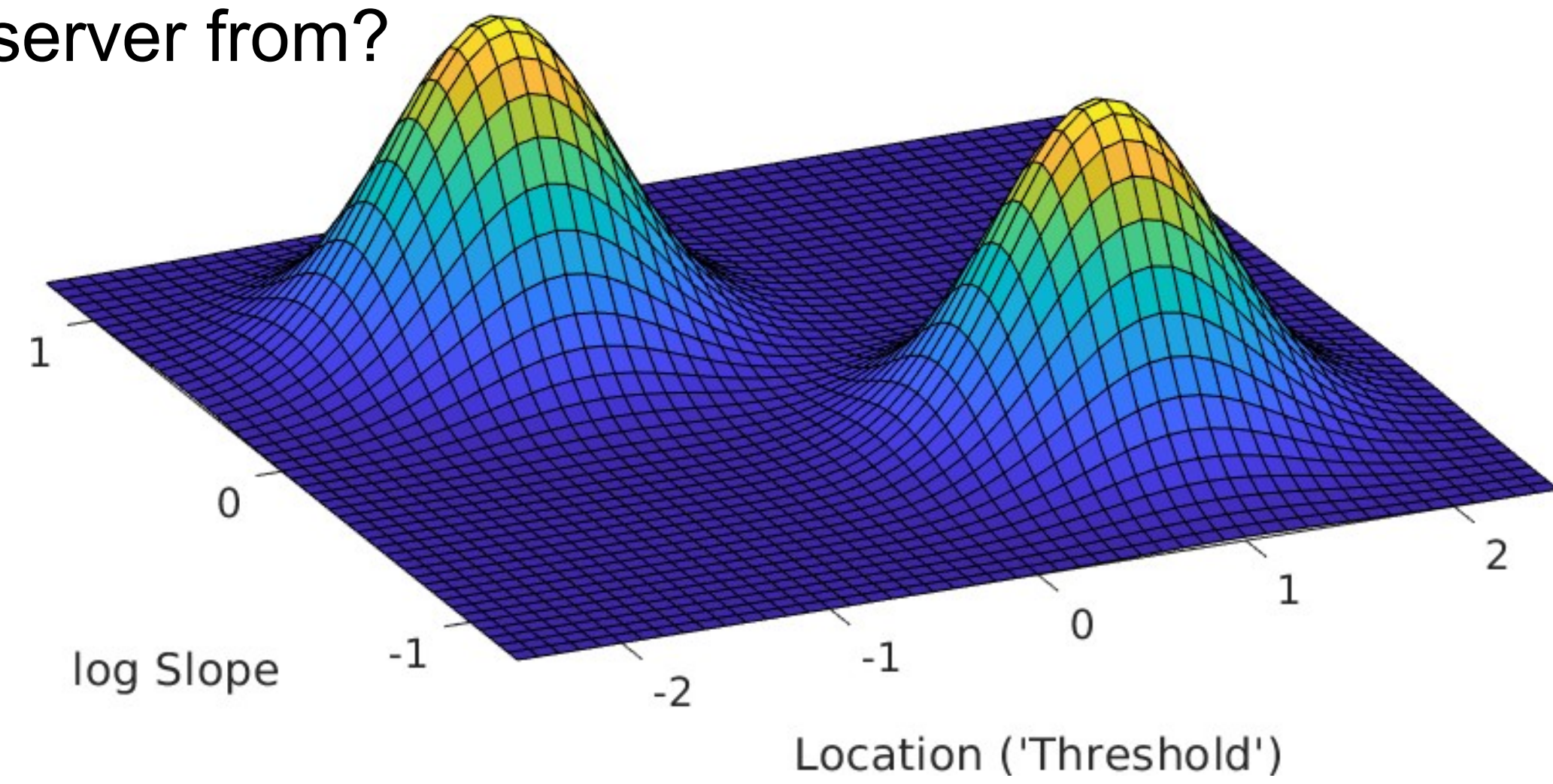
# Optimizing the classification of observers into distinct and diverse categories



Nicolaas Prins  
University of Mississippi

## Introduction

Which of these two populations ('categories')<sup>1</sup> is your observer from?



Here, two adaptive methods to classify observers into distinct and diverse populations are compared. Both maintain a posterior distribution across location ('threshold'), slope, guess rate, lapse rate and population category. The methods have identical priors (i.e., are equally 'informed' of the composition of the populations). Each updates the posterior distribution after each response, then selects the stimulus placement for next trial that minimizes expected entropy (e.g., Kontsevich & Tyler, 1999). The only difference: one method aims to minimize entropy across entire posterior, the other minimizes entropy with respect to population category membership only.

Manipulated here:

- The two populations differ in location, slope, or both.
  - Simulated observer is sampled from population A or B
- Fixed here<sup>2</sup>:
- Bivariate normal populations
  - Cumulative Normal
  - Guess rate (0.25)
  - Lapse rate (0.02)

## Conclusion

Classification of observers into one of multiple distinct and diverse populations can be made significantly more efficient by targeting optimization of categorization exclusively.

## Code:

[www.palamedestoolbox.org](http://www.palamedestoolbox.org)



Look for PAL\_AMPM\_Classify\_Demo.m in PalamedesDemos folder.

## This poster:

[www.palamedestoolbox.org/VSS2022Prins.html](http://www.palamedestoolbox.org/VSS2022Prins.html)



<sup>1</sup>Or: any one of any number of populations  
<sup>2</sup>Many other options exist

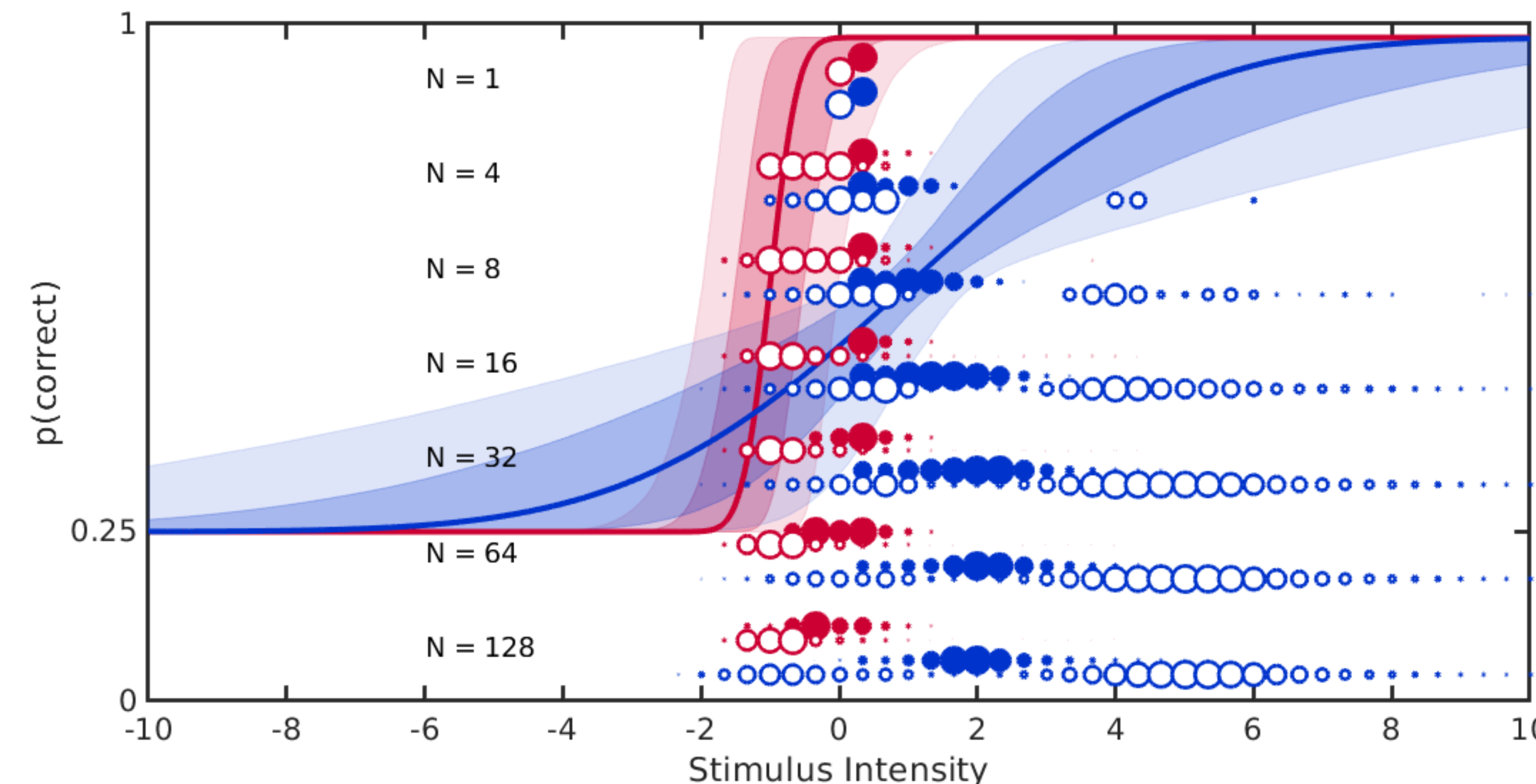
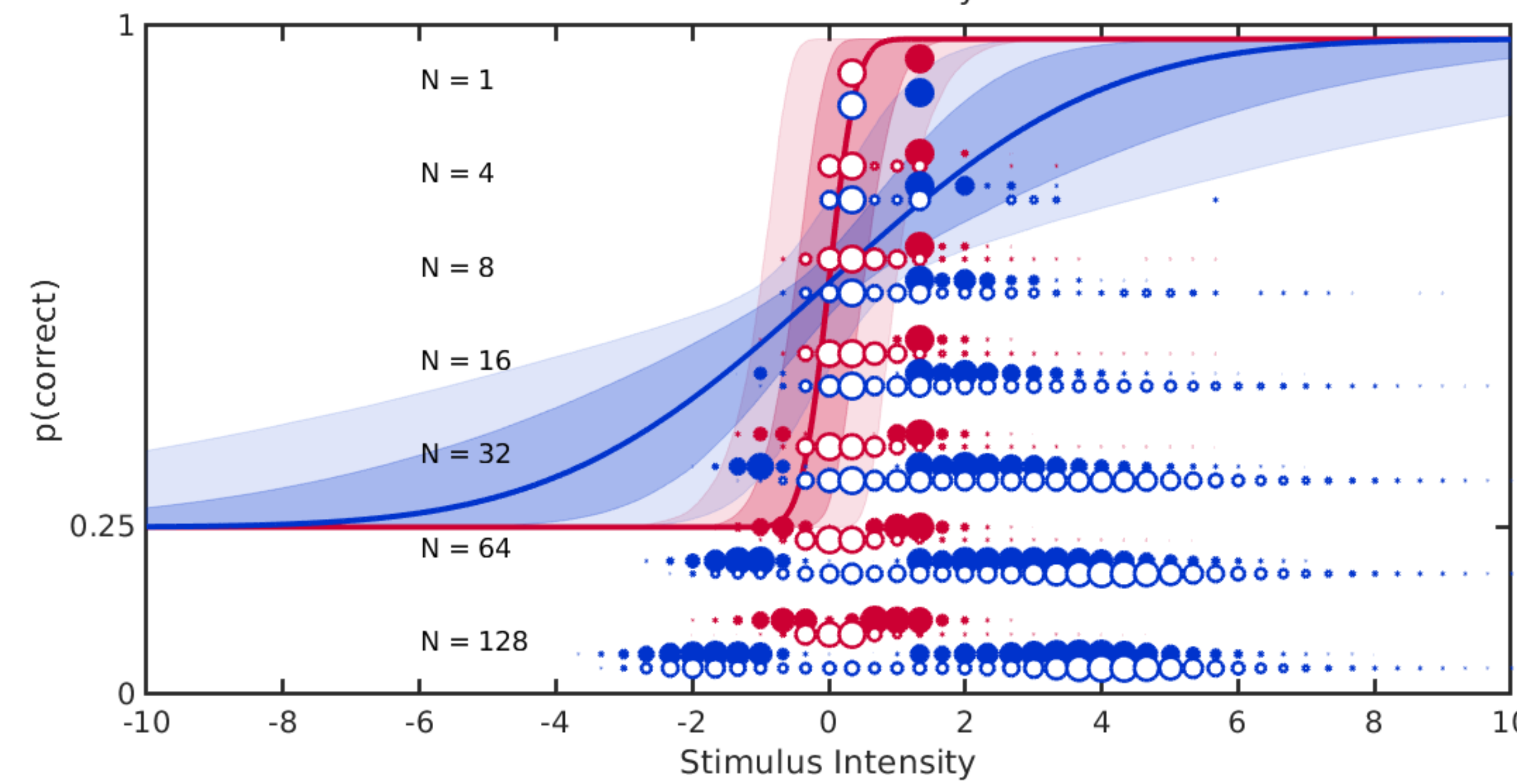
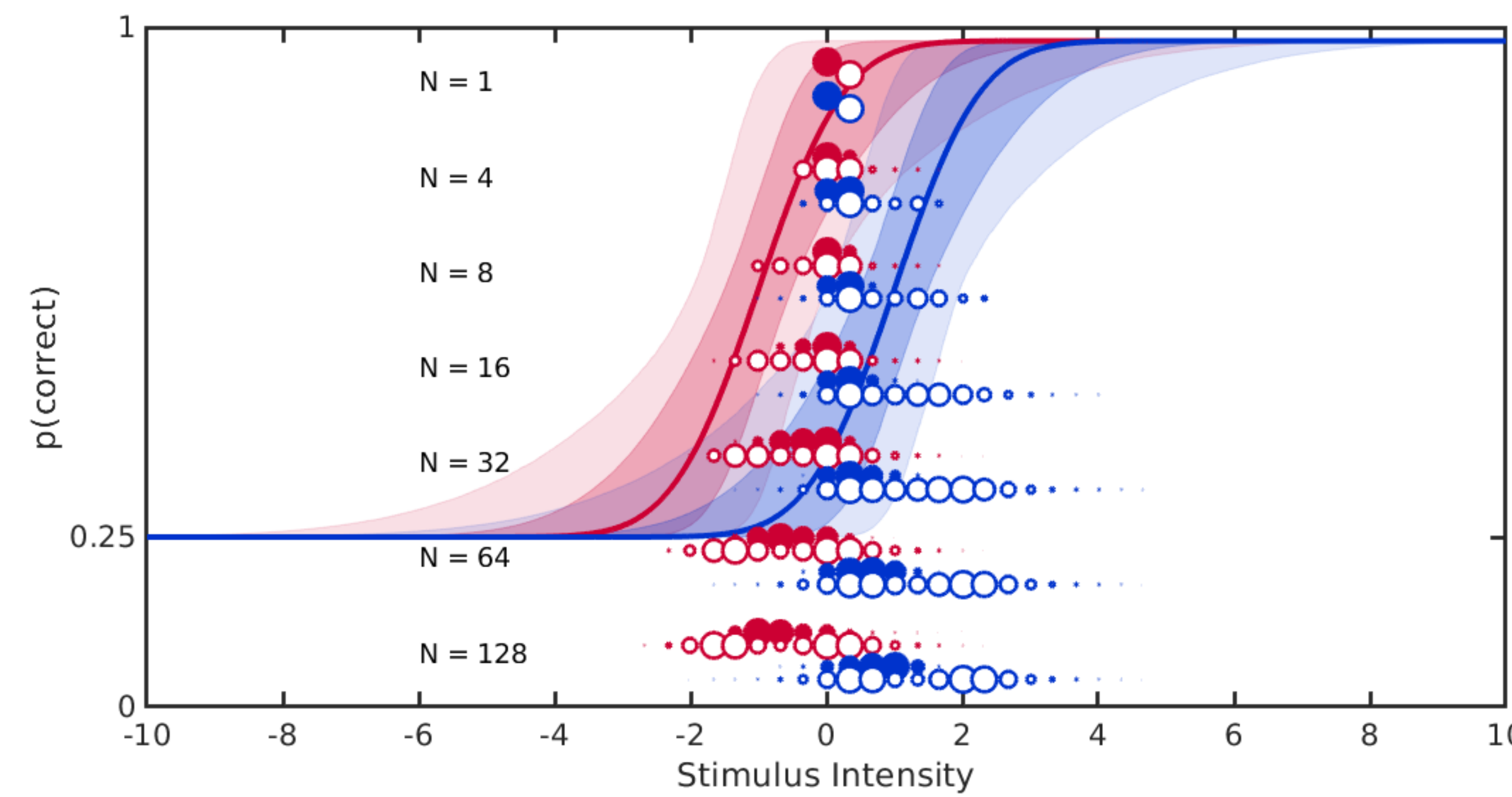
VSS 2022 23.309

References:  
Kontsevich, L.L. & Tyler, C.W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39, 2729-2737.

## Results

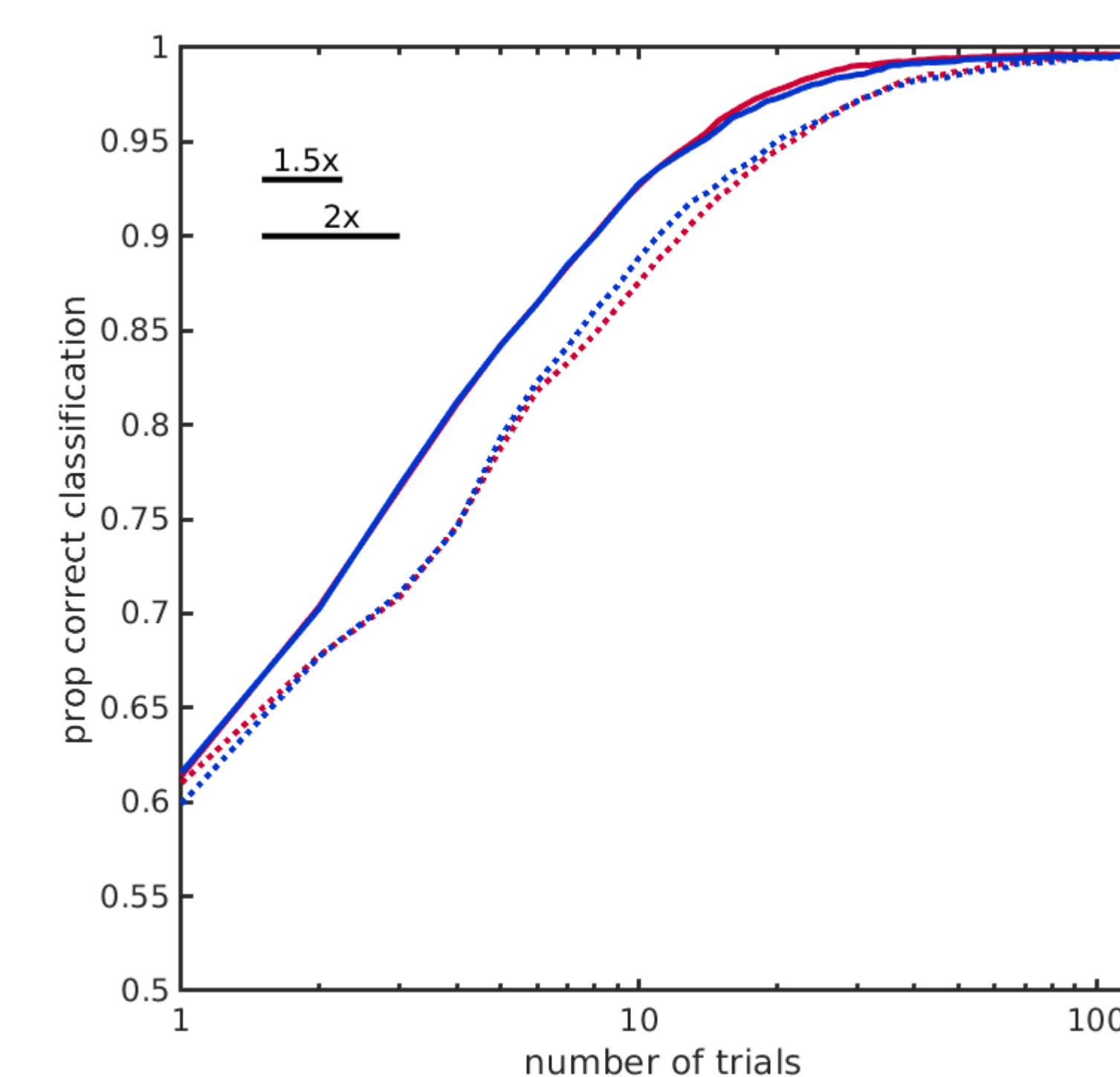
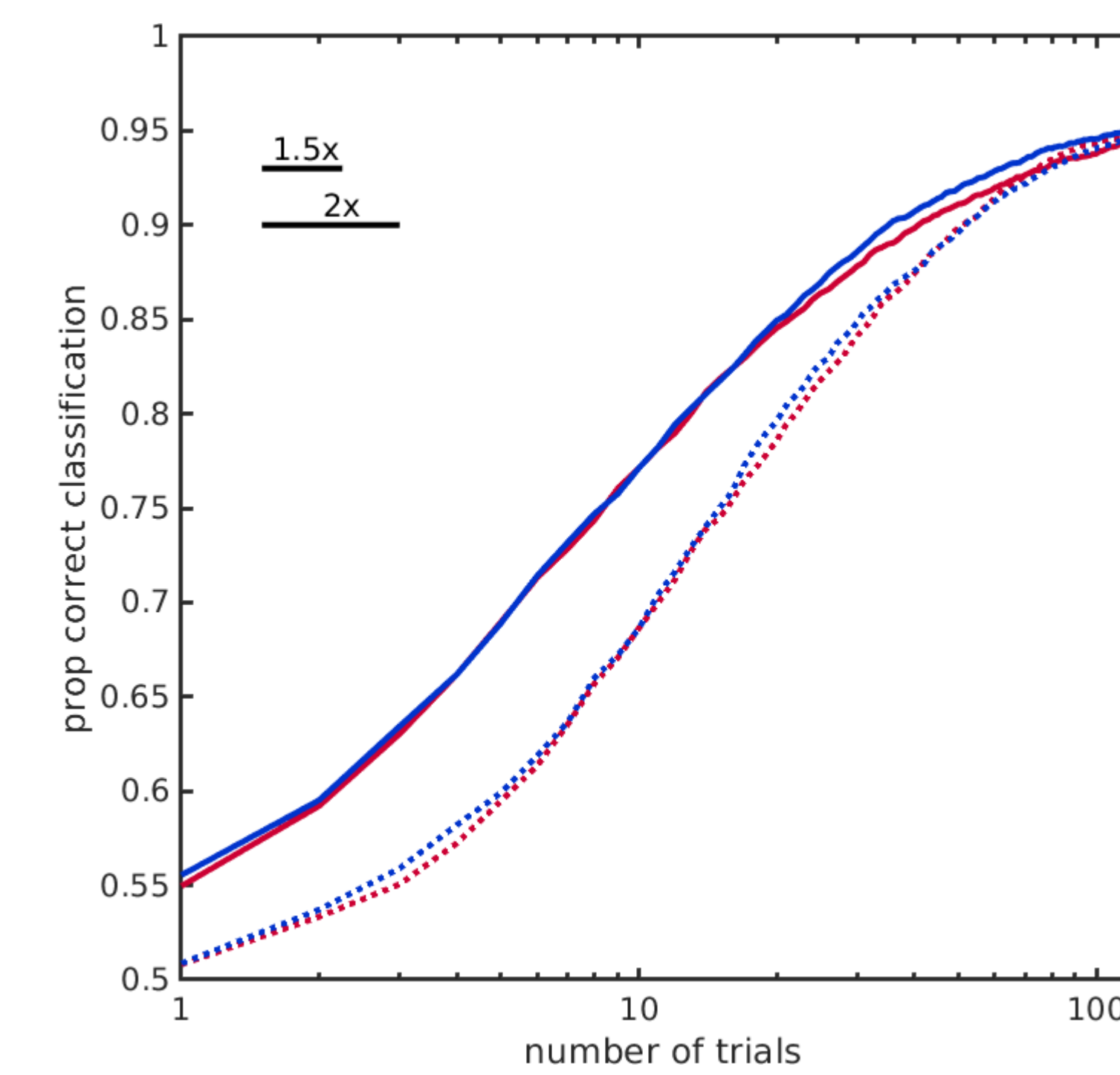
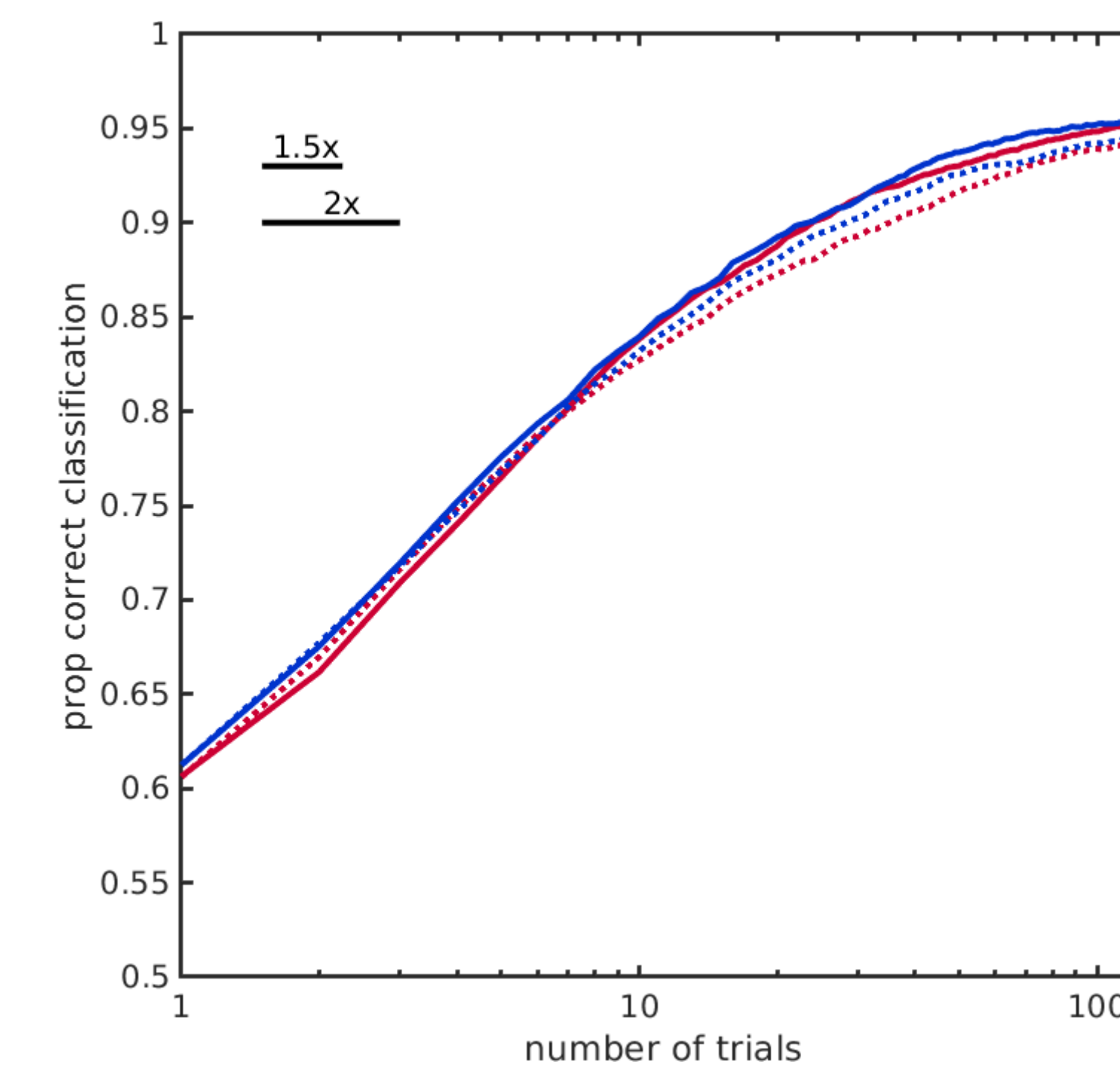
### Placement

(for purposes of highlighting strategy, for this figure only, the generating function was always the mode in population distribution)

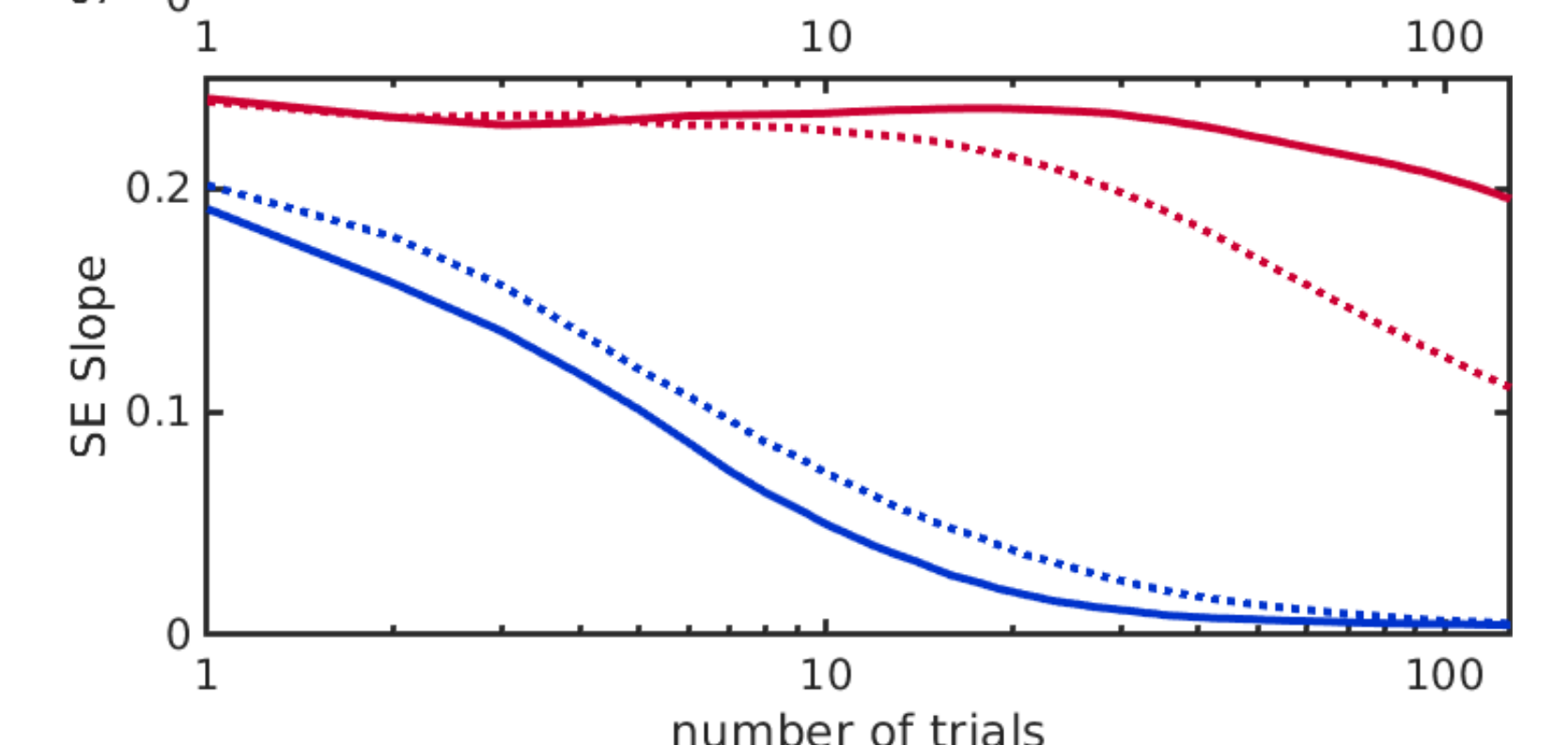
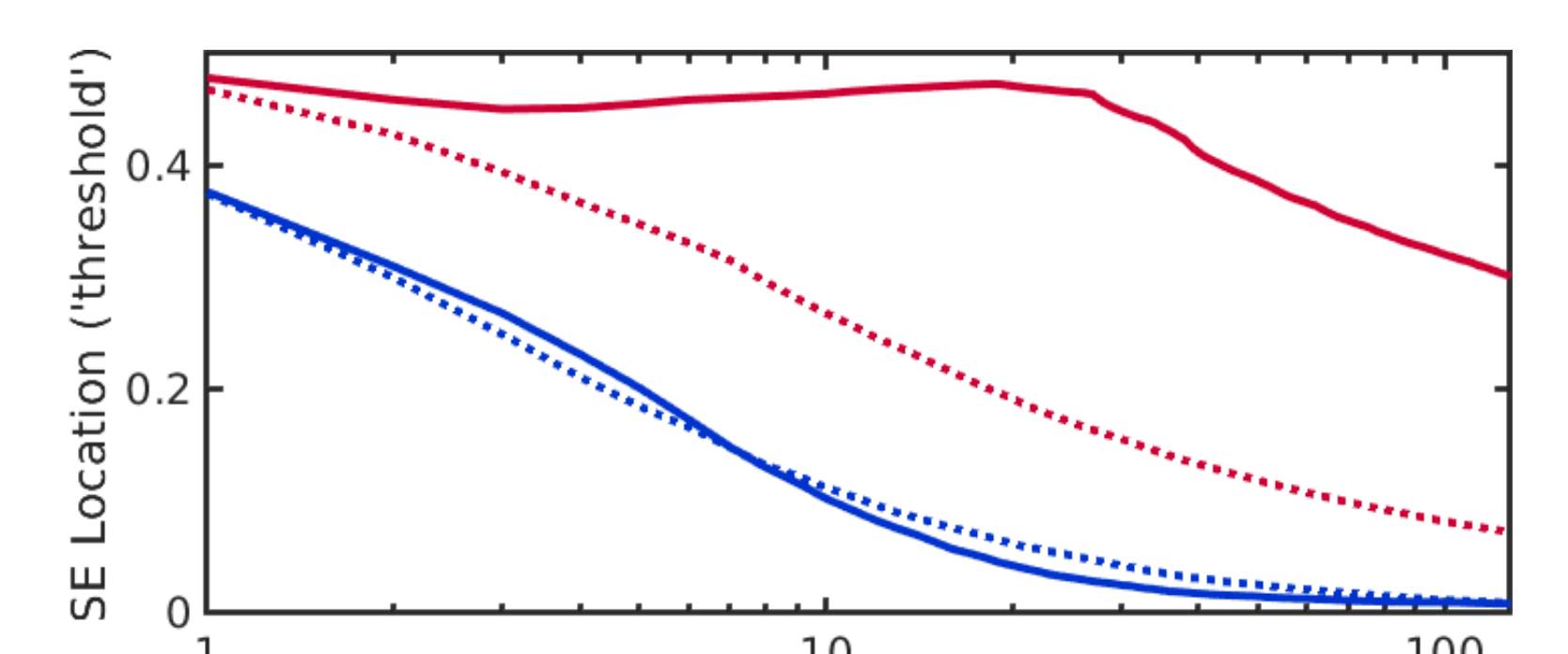
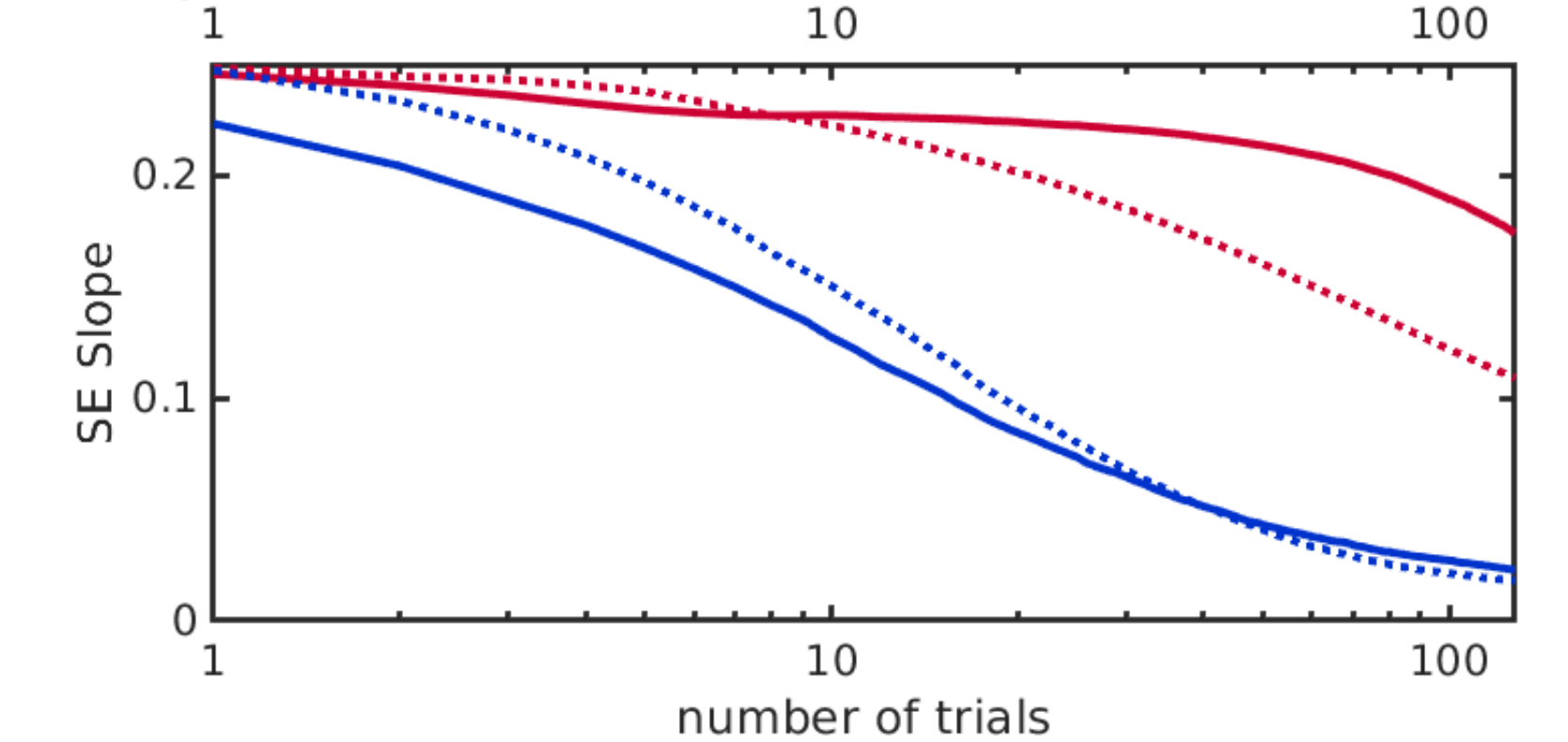
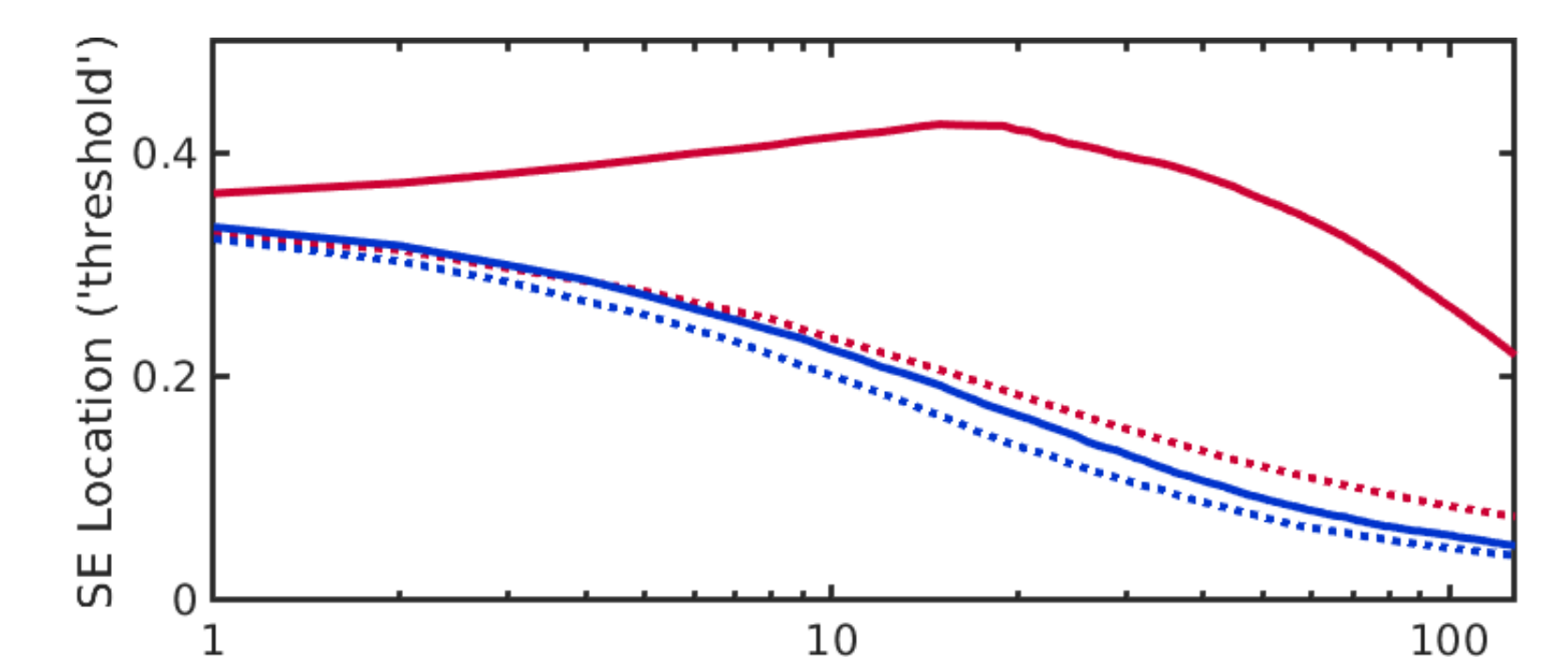
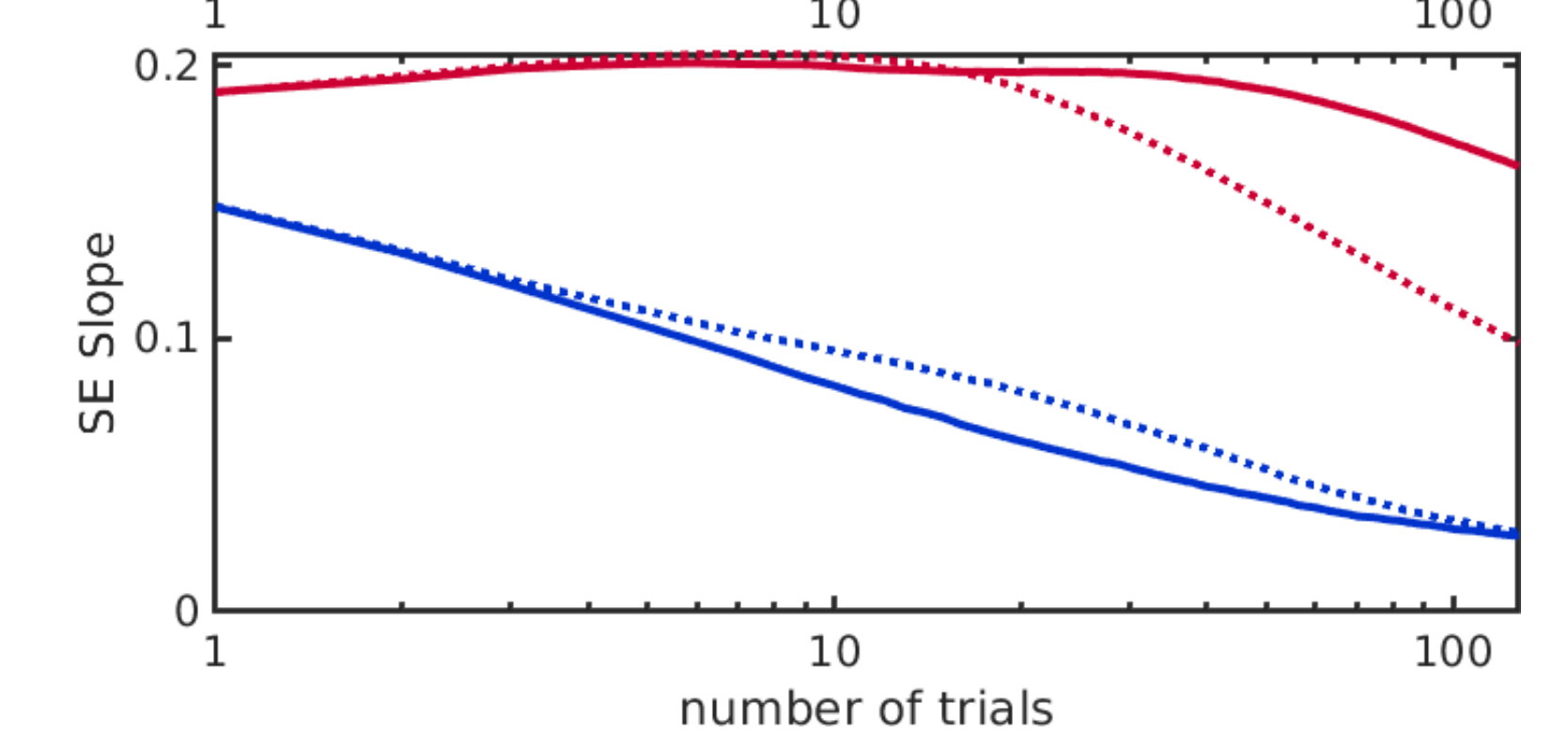
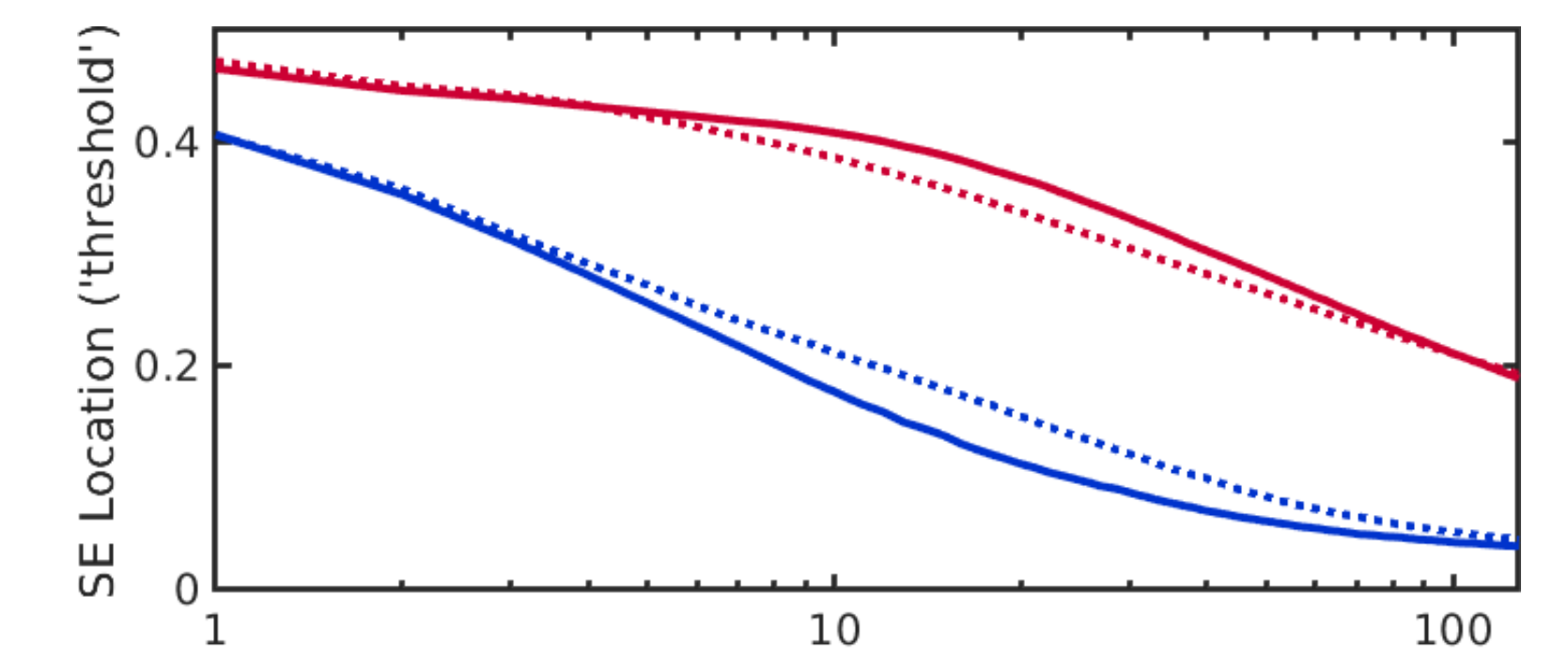


The two populations are displayed here by mode (solid line) and 68% and 95% high-density regions. Symbols indicate placement of trials. Area of symbol proportional to proportion of trials. Color of symbol corresponds to category membership of the simulated observer. Filled: Optimize category membership only, open: optimize all free parameters (location, slope, and category).

### The Gain: Classification Performance



### The Cost: Standard Errors of Parameters



Color of line: Category membership of simulated observer  
Solid: Optimize category membership only, broken: Optimize everything.