Looking into the past: Eye-tracking mental simulation in physical Inference

Ari Beller[1], Yingchen Xu[2], Scott Linderman[1], Tobias Gerstenberg[1]

[1]Department of Psychology, Stanford University

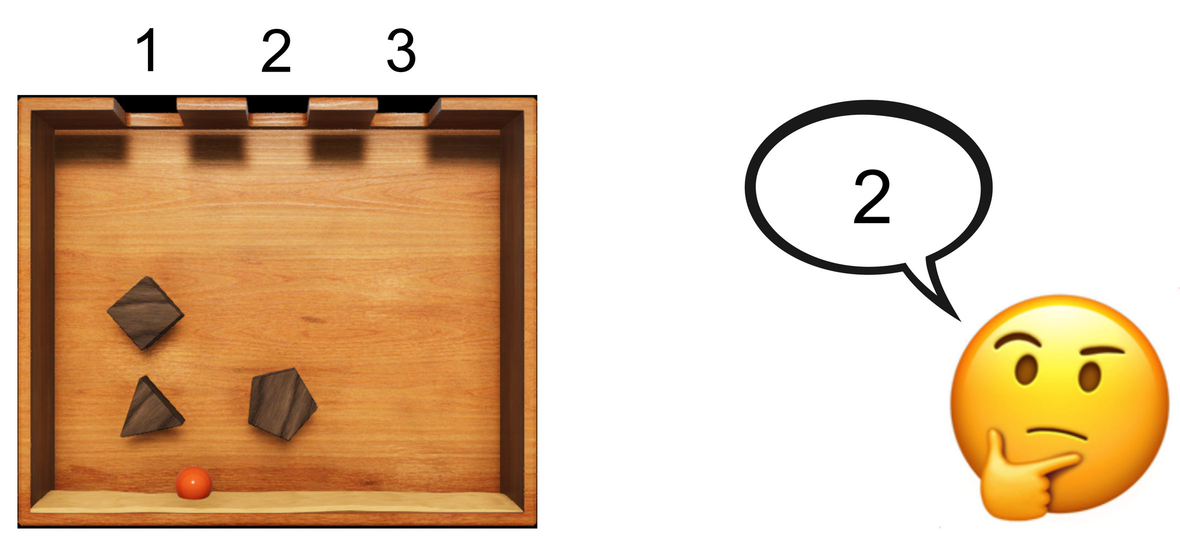
[2]Department of Computer Science, University College London

Diagnostic Inference

- How do people figure out what happened in the past?
- Prior work suggests people do so by running **mental simulations** on **intuitive theories** of the domain (Gerstenberg and Tenenbaum 2017, Battaglia et al. 2013).
- We develop a computational model of the underlying cognitive processes that support causal inference.

Inference Task

In which hole was the marble dropped?

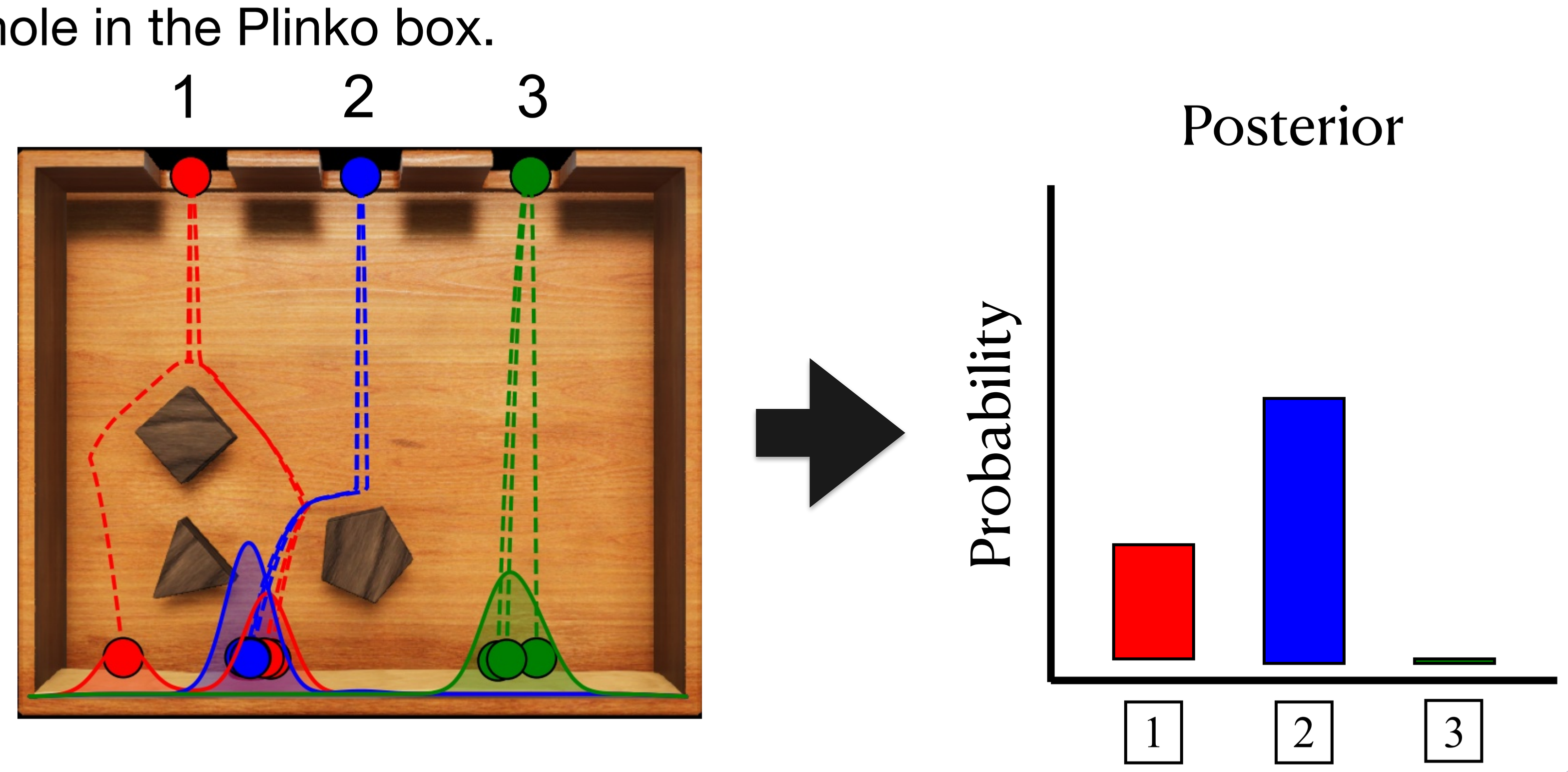


Data

Judgments | Response Time | Eye-Tracking

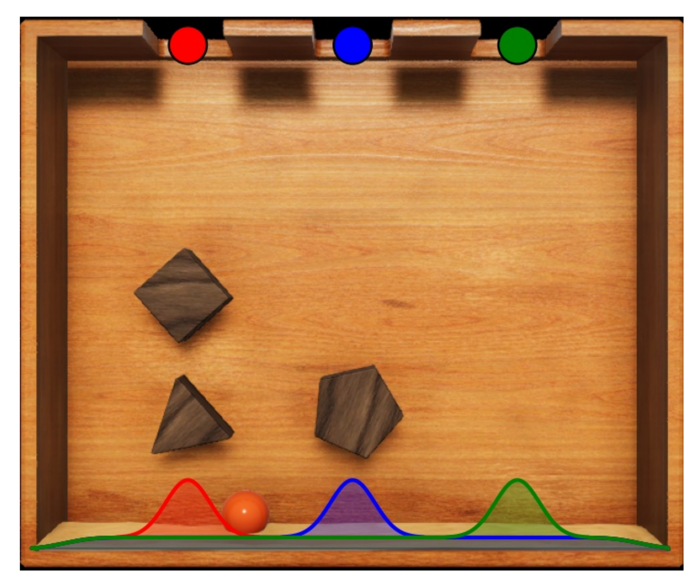
Uniform Sampling Model

Gerstenberg et al (2021) modeled inference in Plinko with a uniform sampling model that ran a fixed number of simulations from each hole in the Plinko box.



Sequential Sampling Model

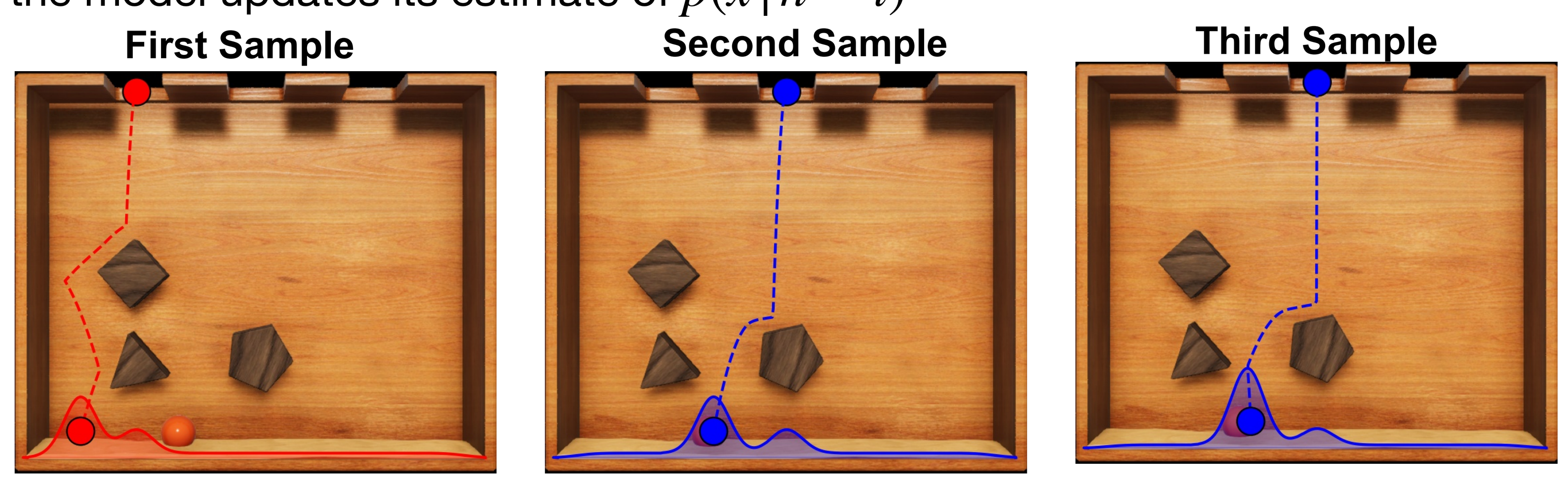
Here we develop a sequential sampler which simulates iteratively until confident enough to judge.



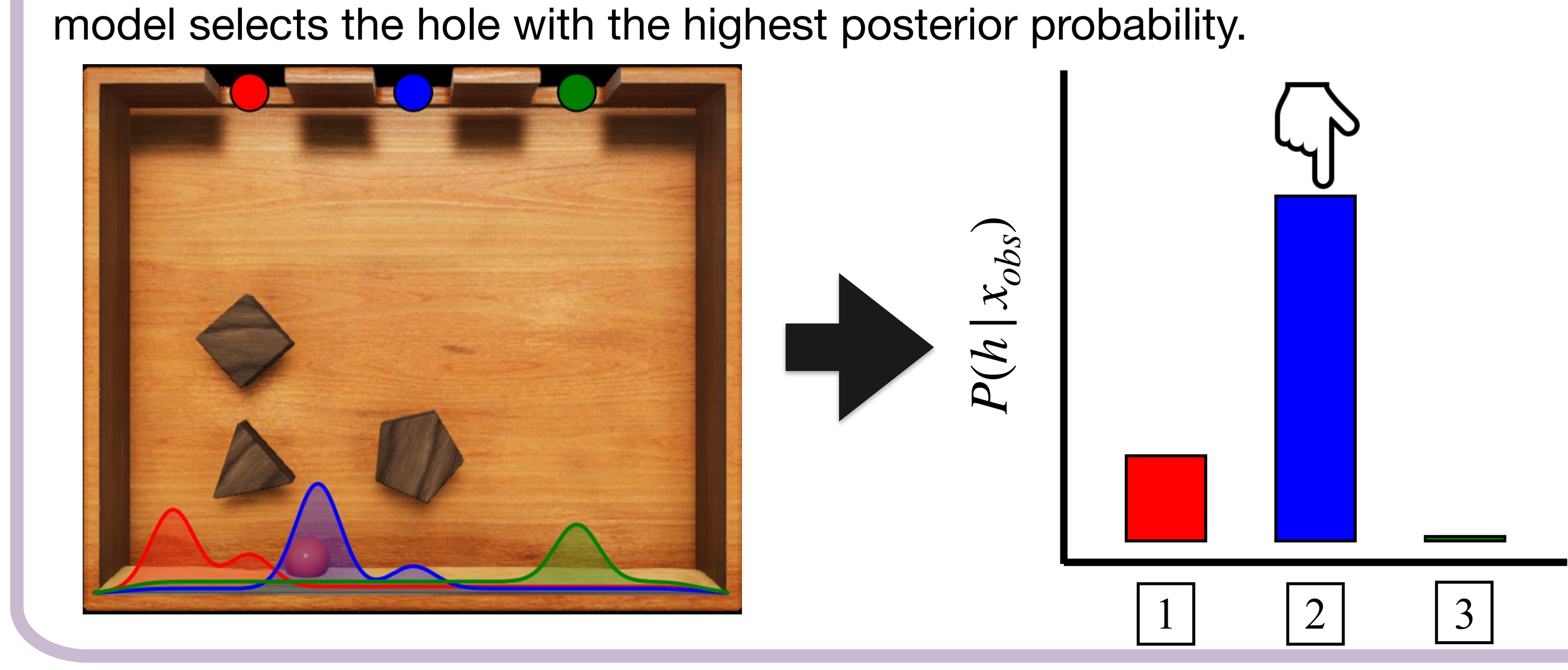
Key
h = hole
i = hole index
x = ball final location
x_{obs} = observed location of the ball

We initialize the model estimate of $p(x\,|\,h)$ with a gaussian prior that expects that the ball will fall near the hole.

The model iteratively selects holes to simulate based on a weighted average of $p(x = x_{obs}\,|\,h = i)$ and the entropy of $p(x\,|\,h = i)$. With each simulation, the model updates its estimate of $p(x\,|\,h = i)$

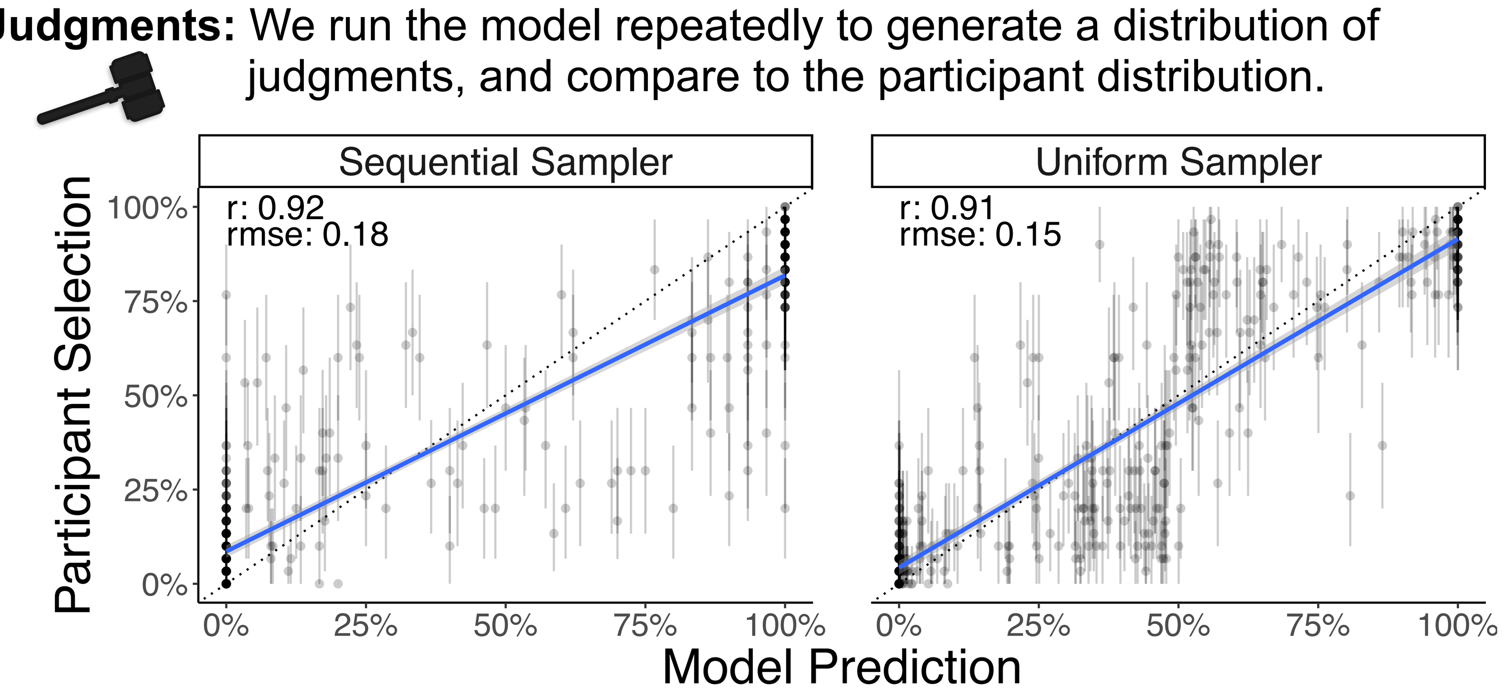


When the entropy of the posterior $p(h\,|\,x_{obs})$ falls below a threshold, the model selects the hole with the highest posterior probability.



Results

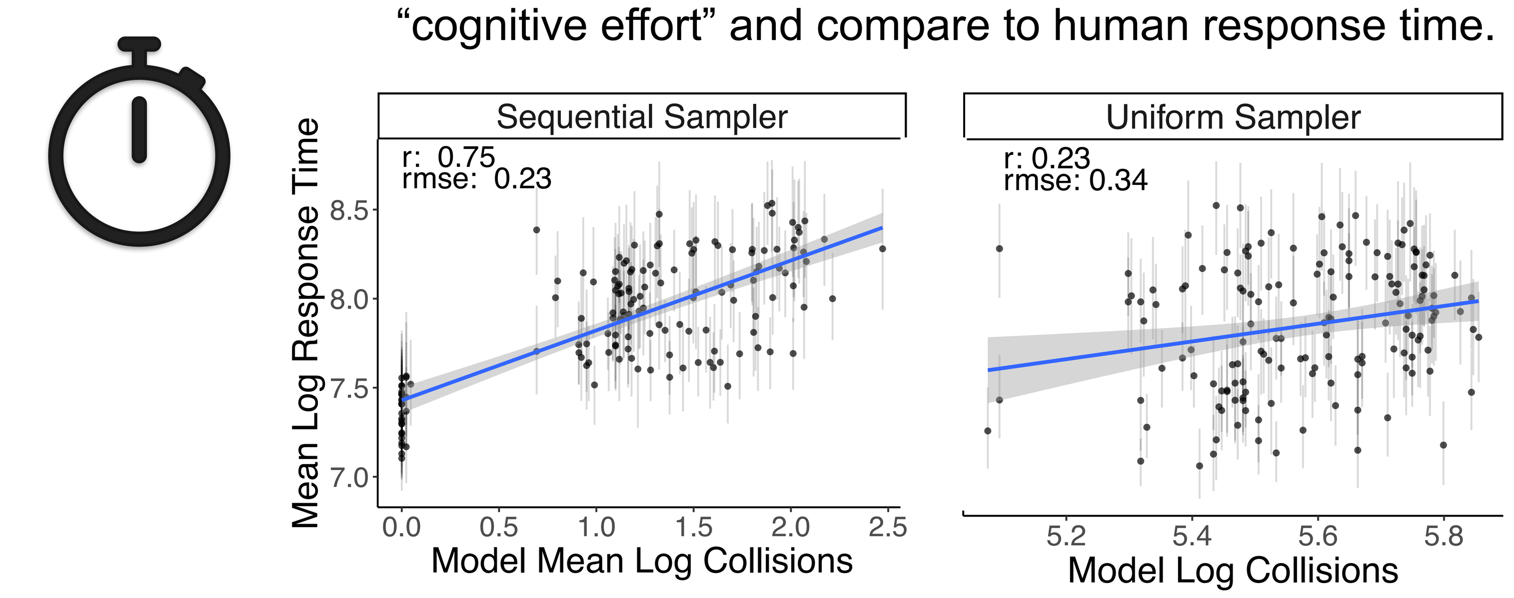
Judgments: We run the model repeatedly to generate a distribution of judgments, and compare to the participant distribution.



Result: Overall both models perform similarly well on the judgment data. We need additional signals to differentiate.

Results (cont.)

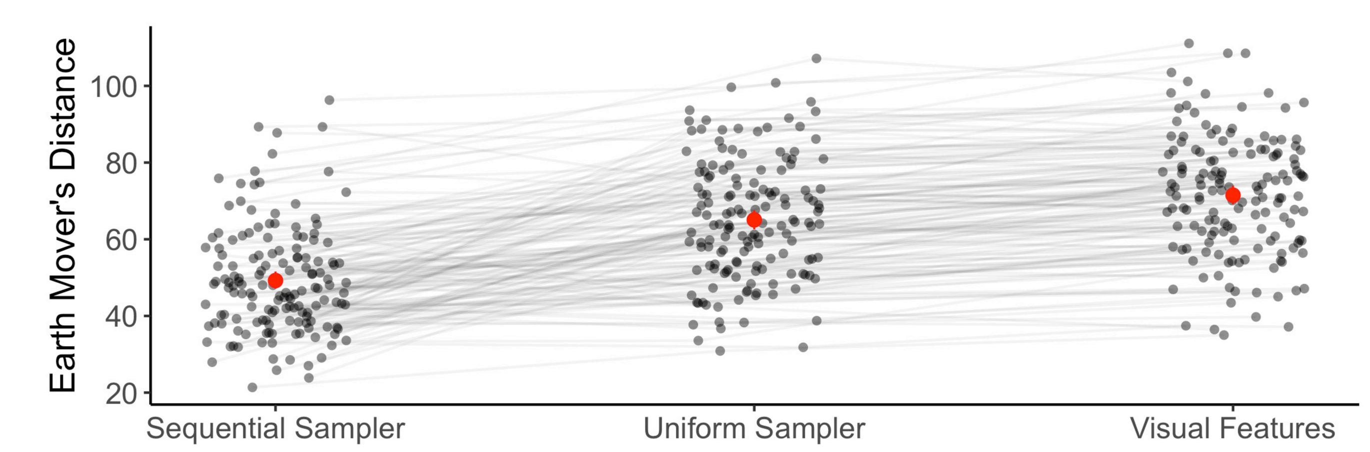
Response Times: We use number of collisions on a trial as an indicator of "cognitive effort" and compare to human response time.



Result: The sequential sampler captures participants' tendency to respond quickly for simple cases and slowly for more complex ones. The uniform sampler cannot capture participants' reaction times as well.

Eye-Tracking: We compute heat maps from features of model behavior and a set of visual features. We predict the distribution of human eye-movement using these feature maps. We compare the difference between the actual and predicted distribution using earth mover's distance.

Result: The sequential sampler only considers plausible hypotheses, just like participants. The uniform sampler also considers hypotheses that participants ignore.



Result: The sequential sampler captures participants' eye-movements best.

Discussion

- We designed a model that explains participants' **judgments, response times,** and **eye-movements** in a novel causal inference task.
- Going forward we'd like to explicitly model how participants use their eye-movements to reduce perceptual and dynamic uncertainty.
- We'd also like to explore how participants' use auditory information to figure out what happened, and how this shows up in the different data signals.

References

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences, 110*(45), 18327-18332.

Gerstenberg, T., Siegel, M. H., & Tenenbaum, J. B. (2021). What happened? Reconstructing the past from vision and sound. https://psyarxiv.com/tfjdk

Gerstenberg, T., & Tenenbaum, J. B. (2017). Intuitive theories. *Oxford handbook of causal reasoning*, 515-548.