The search for useful information in the world is guided by the hypotheses a learner wants to test (e.g., when a doctor orders a diagnostic test based on a potential disease). This external search process is commonly shown to be biased (e.g., the positive test strategy, Klayman & Ha 1987).

One explanation for biased information search is that people have a limited capacity to generate alternative hypotheses (Mynatt et al. 1993). However, prior work has focused on tasks where ongoing generation of multiple hypotheses (and an internal limit on that process) is difficult to measure during information search.

The Rectangle Search Game

OBJECTIVE: With the fewest number of observations possible, learn the shape and location of three rectangles hidden in a 10x10 grid:

SAMPLING PHASE
Participants learn about the hidden shapes by "painting" information at any location in the grid, revealing either an empty square (a "MISS") or part of a solid rectangle (a "HIT").

For every observation, 5 points are added to the score (goal is to get lowest score possible).

TEST ("PAINTING") PHASE
When the participants' score hovers above the shape and location of all three hidden shapes, they can choose to stop searching.

After sampling is over, the shape and location of all three hidden shapes can be revealed. For every square they failed to locate, 2 points are added to the score.

How does capacity-limited hypothesis generation lead to biased information search?

GOALS OF THE STUDY
Using behavior in a spatial search task, we model how an intrinsic limit on the capacity to represent alternatives affects information sampling decisions.

- We measure information search in a task with a well-defined hypothesis space and large number of alternatives.
- Use an ideal observer model to characterize suboptimal sampling decisions.
- Develop a model of the hypothesis generation process that explains the effects of limited capacity on information sampling decisions.

EFFICIENT SEARCH IN "PATCHES" OF INFORMATION
1. After the learner discovers part of a hidden rectangle, the model predicts that the most informative observations are clustered in a surrounding "patch".
   - Observations within a patch are classified as "EXPLORE" trials, while observations outside patches are classified as "EXPLOIT" trials.
   - Overall, participants are more efficient during EXPLORE trials than EXPLOIT trials. In half of the EXPLORE trials, participants chose one of the top 5 ranked samples according to the ideal model (in contrast to only 16% of EXPLOIT trials, revealed in top 5).

FREQUENTLY EARLY STOPPING ERRORS
Another measure of how participants diverged from the ideal model is their tendency to make the following errors:
1. Learning a patch early: the participantdecides to stop sampling within a "patch" of information before uncertainty about that shape was resolved.
2. Stopping the sampling phase early: the participantdecides to stop sampling even though the ideal model predicts that there are informative observations remaining.

Ideal observer analysis

How does people’s search behavior compare to an ideal observer that has perfect knowledge of the hypothesis space (i.e., the set of all possible gameboards)?

1. Ideal observer model represents the full set of gameboards available in the task and optimally updates its belief following each observation using Bayes’ rule.
2. The value of a future observation is measured by the degree to which hypotheses disagree on the predicted outcome of that observation.
3. Since observations in SAMPLING phase are less costly than errors during PAINTING phase, will continue sampling until there is no longer uncertainty about the true gameboard.

Hypothesis generation model

Can a limited capacity to generate alternative hypotheses account for suboptimal information search?

- SEQUENTIAL HYPOTHESIS SEARCH
  - Model represents small number of hypotheses about one of the three target shapes.
  - The capacity (M) is the number of "patches" of units that are divided among competing hypotheses.
  - Hypothesis set is updated following new observation, assuming that inconsistent hypotheses are discarded.
  - New hypotheses are generated by making "proposals" based on the previous hypothesis set, leading to an incremental, real process of hypothesis generation.

- SAMPLING BY UNCERTAINTY
  - Value of a future observation determined by disagreement between current hypothesis set.
  - If only a single hypothesis is represented (M=1), the model employs a "positive less strategy" by sampling at observed locations that have the least disagreement.

As capacity of the model increases, greater number of possible hypotheses contributes to the measurement of uncertainty, making the sampling decision in favor of more informative observations.

CONCLUSIONS
- Found systematic divergence between participants' behavior and the ideal learner:
  1. People are more efficient when expediting patches of information related to a single target.
  2. People frequently leave a patch or stop sampling altogether before reducing uncertainty completely.
- Across 5 participants, a limited capacity hypothesis generation model better predicts search behavior than the ideal observer due to its limited representation and sequential process of switching between target shapes.
- Performance was best captured by a hypothesis generation model with an intermediate capacity (M=10) as compared to a low-capacity (M=2) and "confirmatory" model (M=1).
- Future work will measure ongoing changes in capacity limit as a result of extrinsic task demands (e.g., working memory load).