The impact of biased hypothesis generation on self-directed learning

Doug Markant (markant@mpib-berlin.mpg.de)

Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin

Center for Adaptive Rationality
Max Planck Institute
for Human Development

Self-directed learning involves an ongoing interaction between active information search and sequential hypothesis testing. The hypotheses that people generate as they learn form the basis for reasoning, prediction, and further exploration.

Previous work suggests that **hypothesis generation may be a critical bottleneck that limits the efficacy of self-directed learning**. Hypothesis generation is often biased by prior knowledge, processing constraints, and failures to consider alternatives.

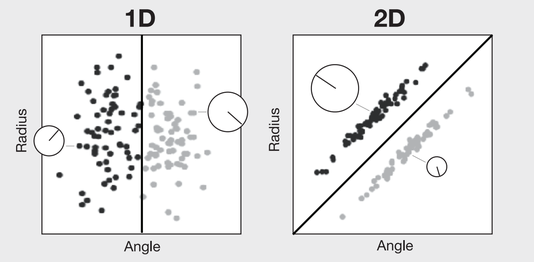
How does biased hypothesis generation affect the ability to acquire categorical rules through self-directed learning?

This study examines how **feature representations bias hypothesis generation during self-directed learning.** Distinct features were predicted to favor generation of 1D hypotheses, whereas matched features were predicted to favor 2D hypotheses that involve integrating feature values.

Self-directed category learning: A gap between 1D and 2D rules

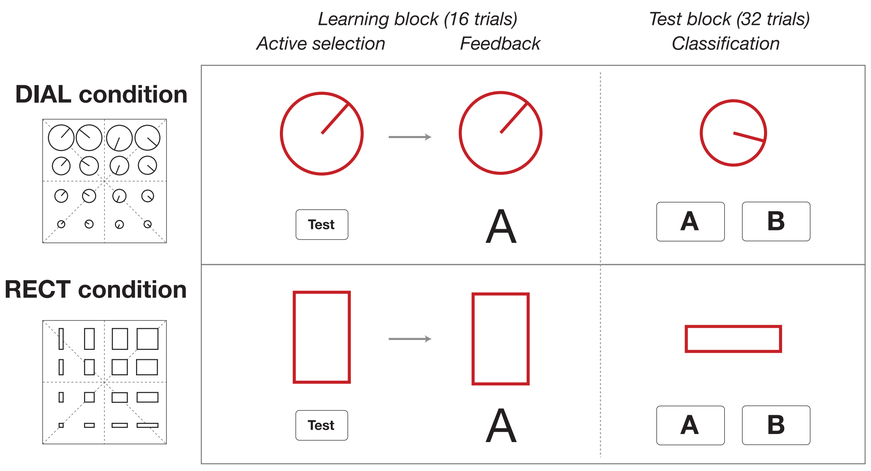
Markant and Gureckis (2014) compared active selection and passive reception in a perceptual category learning task involving either 1D or 2D rules. For 1D rules, selection-based learners outperformed reception-based learners and selected stimuli close to the true category boundary. For 2D rules, however, there were no differences between selection- and reception-based learning. Participants in the 2D condition classified test items according to simpler 1D hypotheses throughout the task.

Why did selection-based learners have a bias for 1D hypotheses? Whereas existing theories (e.g., COVIS) assume an intrinsic limit on the complexity of hypotheses that can be generated, this bias may have been driven by the use of perceptually distinct features that favored the generation of 1D hypotheses, rather than 2D hypotheses based on the relative magnitudes of the two features.



Experiment 1

GOAL: Learn which shapes belong to category A or category B



Experiment 2

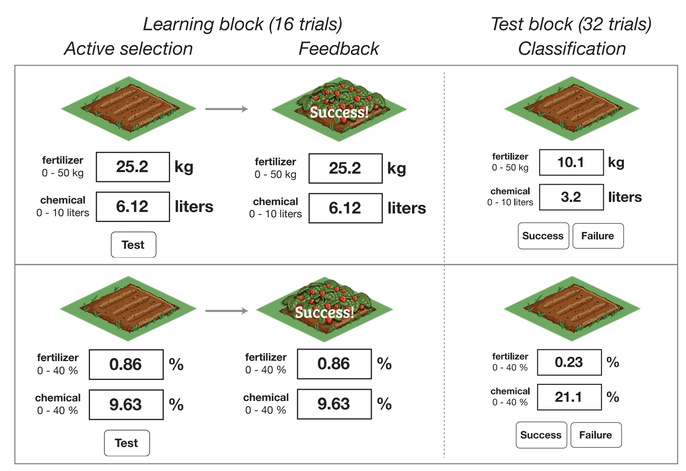
GOAL: Learn which combinations of substances produce a successful or failed crop in a virtual farming game

ABS condition

"Absolute" features with different ranges and described in different units (kg or liters)

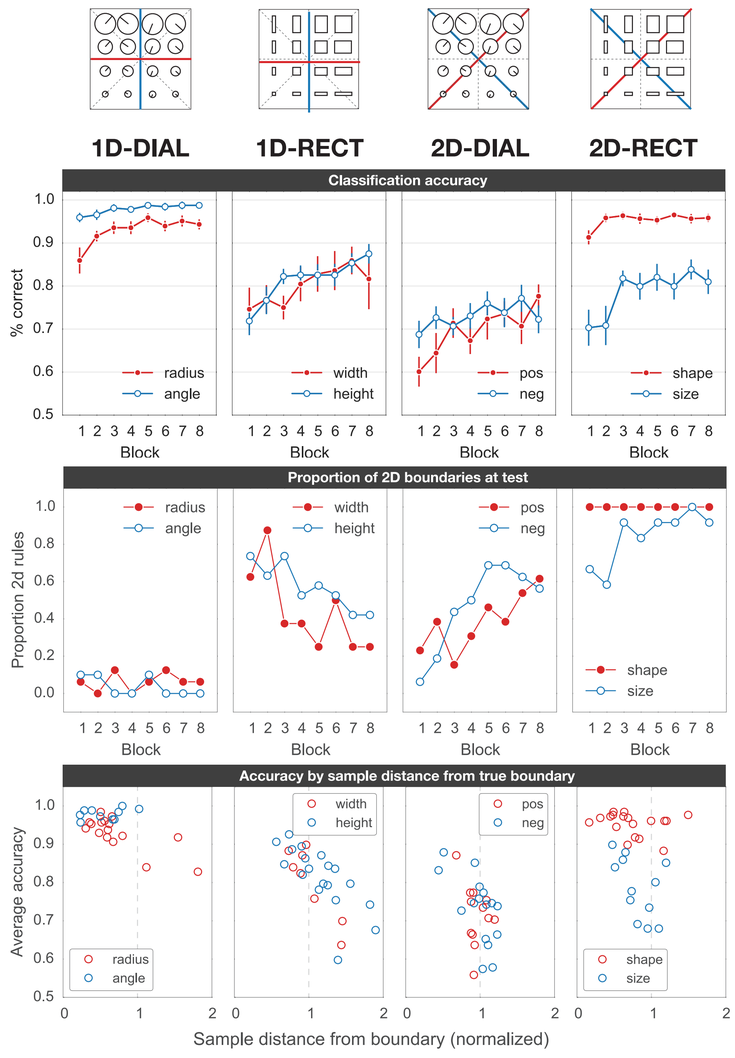
REL condition

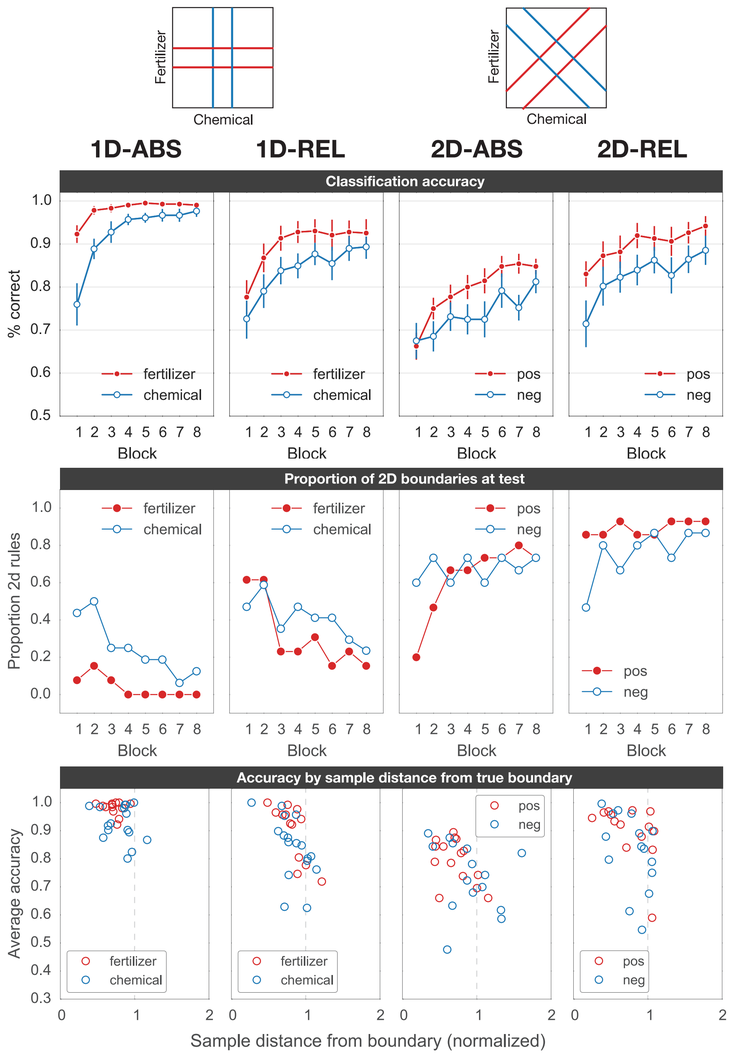
"Relative" features with a common range and described in the same units (%)



Design

- All participants learned through self-directed selection of training items, and were assigned to learn either a 1D or 2D rule.
- On each training trial, they selected a stimulus and received feedback about its category label (see right).
- Classification tests were based on a grid of items uniformly distributed over stimulus space
- 8 training and test blocks (interleaved)





Results

- Replicated gap between 1D and 2D accuracy reported by Markant & Gureckis (2014) for both DIAL stimuli (Exp. 1) and ABS stimuli (Exp. 2).
- Compared to distinct feature representations (DIAL and ABS conditions), matched features led to faster learning of 2D rules but slower learning of 1D rules, in both Exp 1 (RECT conditions) and Exp. 2 (REL conditions).
- Classification responses were modeled with Bayesian logistic regression to find the proportion of blocks best-described by 2D response boundaries. In both Exp. 1 and Exp. 2, matched features led to increase in the proportion of 2D boundaries.
- As observed by Markant & Gureckis (2014), accuracy was negatively correlated with the average distance of participants' selections from the true boundary (with the exception of 1D-DIAL and 2D-RECT conditions in which performance was at ceiling).

Conclusions

- Despite the ability to control the selection of training data, performance depended on whether the feature representation favored the generation of hypotheses consistent with the target rule (i.e., RECT and REL conditions improved learning of 2D rules but *impaired* learning of simpler 1D rules).
- Given self-directed control and a matched feature representation, participants were relatively efficient at learning 2D rules (in contrast to previous evidence of poor learning under passive conditions and highly distinct feature representations).
- These results highlight how hypothesis generation is biased by properties of the learning environment, including the relative salience of individual features and higher-order relations. The efficacy of self-directed learning depends not only on the ability to test hypotheses through information search, but on whether the environment facilitates the generation of hypotheses consistent with a target concept.

Markant, D. and Gureckis, T. M. (2014). Is it better to select or receive? Learning via active and passive hypothesis testing, JEP:G, 143(1):94-122.