

Science as an Efficient Game of Mastermind

Michael H. Birnbaum

California State University, Fullerton

September, 2023

Mastermind Game- Basic Game



Auf Deutsch => “SuperHirn”



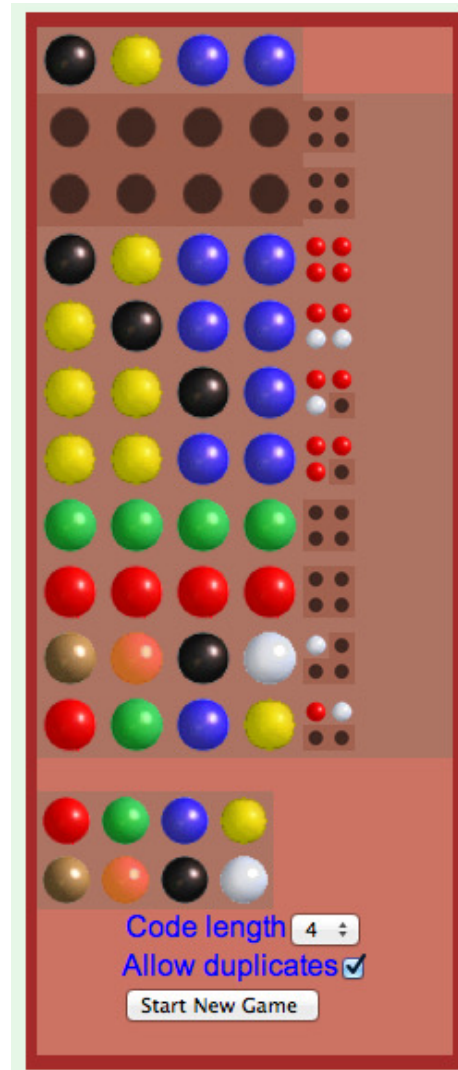
Mastermind Game

- Goal to find secret code of colors in positions. In “basic” game, there are 4 positions and 6 colors, so, $6^4 = 1296$ hypotheses. 8 colors: $8^4 = 4096$
- Each “play” of the game is an experiment that yields feedback related to the accuracy of an hypothesis.
- For each “play”, feedback = 1 colored peg for each color in correct position and 1 white peg for each correct color in wrong position. Results are published. Design new experiments.

Play Mastermind Online

- <http://www.web-games-online.com/mastermind/index.php>
- <http://www.archimedes-lab.org/mastermind.html>
- (Mastermind is a variant of “Bulls and Cows”, an earlier code-finding game. *Word Mastermind*; recent version: *Wordly*.)

A Game of Mastermind- $4,096 = 8^4$



Analogies

- EXPERIMENTS yield results, from which we revise our theories.
- RECORD of experiments and results is preserved.
- Experiments REDUCE THE SPACE of theories compatible with evidence.
- Hypotheses can be PARTITIONED with respect to components.

Science vs. Mastermind

- In Mastermind, feedback is 100% accurate; in science, feedback contains “error” and “bias.” Repeat/revise the “same” experiment, different results.
- In Mastermind, we can specify the space of hypotheses exactly, but in science, the set of theories under contention expands as people construct new theories.
- In Mastermind, we know when we are done; science is never done.

Analogies

- EFFICIENT Mastermind is the goal: Find the secret code with fewest experiments.
- If FEEDBACK IS NOT PERFECT, results are fallible, and it would be a mistake to build theory on such fallible results.
- REPLICATION is needed in empirical research, despite the seeming loss of efficiency.

Hypothesis Testing vs. Mastermind

- Suppose we simply tested hypotheses, one at a time and a significance tests says “reject” or “retain”?
- With 1296 hypotheses, we get closer to truth with each rejection--BARELY.
- Now suppose that 50% of the time we fail to reject false theories and 5% of the time we reject a true theory. Very slow progress.
- Clearly, significance testing this way is not efficient. More INFORMATIVE FEEDBACK needed.

Experiments that Divide the Space of Hypotheses in Half

- Basic game = 1296 Hypotheses
- Suppose each experiment cuts space in half:
1296, 648, 324, 162, 81, 40.5, 20.25, 10.1, 5.1, 2.5, 1.3, done. 11 moves.
- But typical game with 1296 ends after 4 or 5 moves, infrequently 6.
- So, Mastermind is more efficient than “halving” of the space.

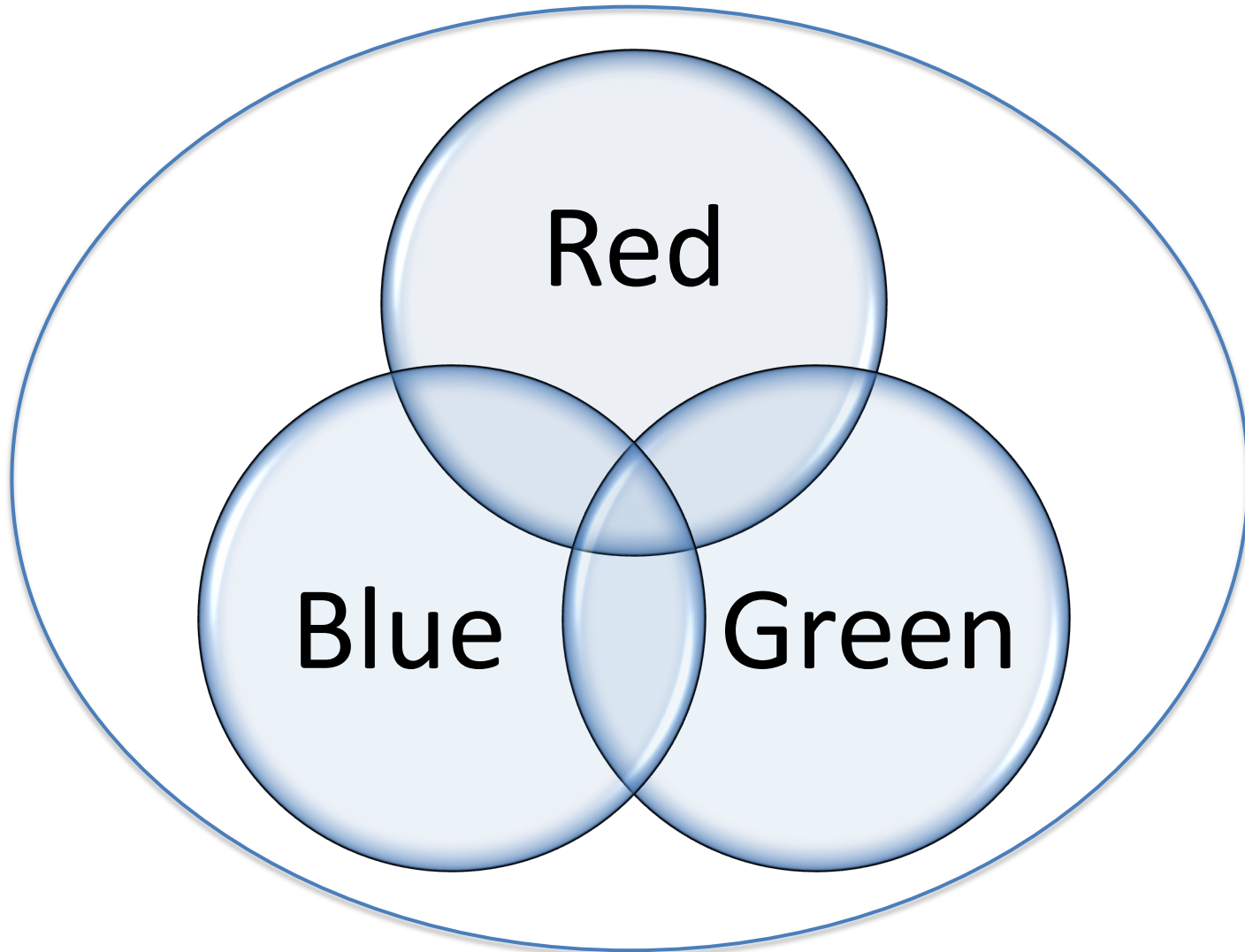
Index of Fit Informative?

- Suppose we assign numbers to each color, $R = 1$, $G = 2$, $B = 3$, etc. and calculate a correlation coefficient between the code and the experimental results?
- This index could be highly misleading, it depends on the coding and experiment.
- Fit can be higher for “worse” theories. (“Devil rides again” papers PB, 1970s).

Psychology vs. Mastermind

- Mastermind: only ONE secret code.
- In Psychology, we allow that different people might have different individual difference parameters.
- Even more complicated: Perhaps different people have different models.
- As if, different experiments in the game have DIFFERENT secret codes.

Partitions of Hypotheses



Testing Critical Properties

- Test properties that do not depend on parameters.
- Such properties partition the space of hypotheses, like the test of all REDs.
- For example: CPT (including EU) implies STOCHASTIC DOMINANCE. This follows for any set of personal parameters (any utility/value function and any prob. weighting function).

Critical Tests are Theorems of One Model that are Violated by Another Model

- This approach has advantages over tests or comparisons of fit.
- It is not the same as “axiom testing.”
- Use model-fitting to rival model to predict where to find violations of theorems deduced from model tested.

Summary

- Study of Alternative Explanations is like a game of Mastermind.
- Experiments designed to inform us as to the space of possible theories that are compatible with the data.
- Test critical properties to distinguish theories.

30 Years Later- Old Bull Story



Summary of Exercises

- Bare bones Web page (memorize).
- Experiment construction using FactorWiz (RB)
- Heider Study: You like Bill; Bill likes John
Edit: add text; change font face, change background color.

Import data to Excel. Use sort or filters to extract data. Use text to columns. Find means. Arrange in factorial design. Use solver to fit models. Multiplicative model fits well.