Function Learning

All of psychology?

Eric Schulz
MPI for Biological Cybernetics

Preamble

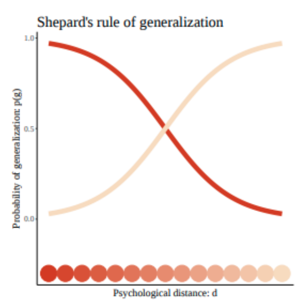
- **When I heard that there's a topic called function learning:**
  - I thought: "What is not a function"?
  - Shouldn't this just be –like– psychology?
- **In reality:**
  - Function learning was quite a narrow topic
  - Relatively exotic and abandoned (still!)
- **Goal: Modernize theories of function learning!**

The original dispute: Shepard

Description

- Roger Shepard's law
- Decay of P(same response)
- Response set is fixed
- Universal law
- Thus: it should always be like that!

Overview

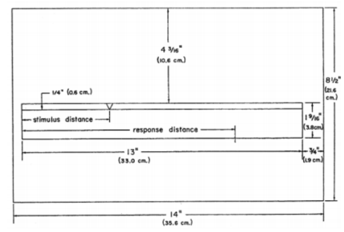


The original dispute: Carroll

Description

- Douglas Carrol's doubts
- Sometimes animal shows different response
- Variance left unexplained
- Function learning like least squares?
- Let's study this!

Overview



What did Carroll find?

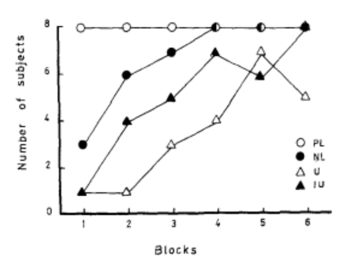
- Participants learned S-R combinations better if these were governed by an underlying function instead of a random mapping
- Linear functions are easier to learn than non-linear functions
- Instead of mapping response to a novel stimulus to response of the closest experienced stimulus, participants were able to extrapolate response
- Modelled by a form of least-square regression
- Rule-based account of function learning: Participants approach function learning with a set of fixed functional rules

Extension by Brehmer

Description

- Berndt Brehmer's rigorous tests
- Positive linear vs. negative linear vs. U-Shaped vs. Inverse U-Shaped
- Linear positive by far the easiest
- Postulated sequential hypothesis tests
- Still rule-based

Overview

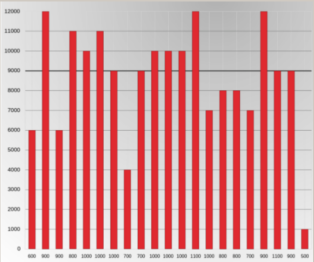


Controlling functions

Description

- Donald Braodbent's sugar factory
- Participants had to learn how a continuous value called "work force" relates to the amount of sugar a factory can produce
- Later produce a given value of sugar, thereby having to control the factory
- Linear easier to control than exponential
- Recommended to "linearize" control

Overview



Participants love linear-positive functions

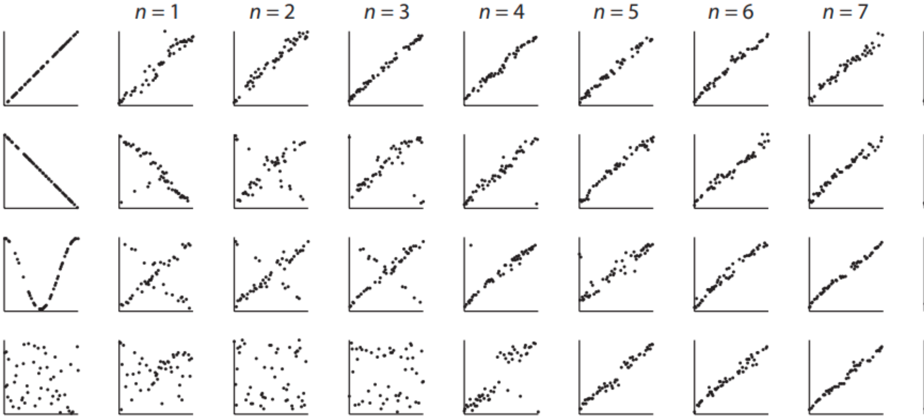


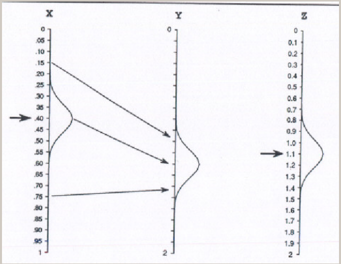
Figure: Iterated learning experiment (Kalish et al.)

Similarity-based function learning

Description

- Jerome Busemeyer used neural networks
- Associative Learning Model (ALM)
- Gaussian similarity functions with generalization gradient
- Can reproduce main effects

Overview



Extrapolation is important

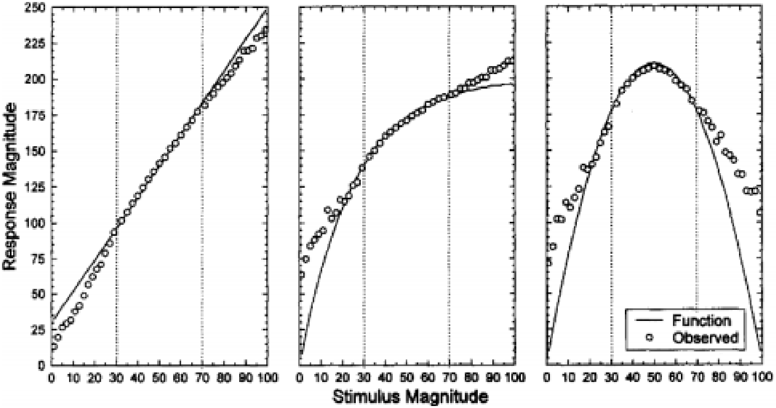


Figure: Extrapolation experiment (DeLosh et al.)

Hybrid models

EXAM

- Extrapolation-Association Model (Busemeyer et al.)
- Similarity-based interpolation, linear rules for extrapolation
- Falls back onto a parametric representation of whenever extrapolation is required
- Does not capture the ability to extrapolate non-linear functions (Bott & Heit, 2004)

POLE

- Population of Linear Experts (Kalish et al.)
- Approximates functions using piece-wise linear representations
- Can explain knowledge partitioning
- Does not capture oder of presentation effect (Byun, 1995)

Function learning: Classic results

- Linear functions are easier to learn than non-linear functions
- Linear functions with positive slope seem to be the default
- Interpolation seems smooth
- Extrapolation has linearity bias
- Participants can partition the input space
- Sequentially ordered presentation facilitates learning
- Learning non-linear functions is possible
- Hybrids of rule-based and similarity-based models work best

Function learning everywhere?

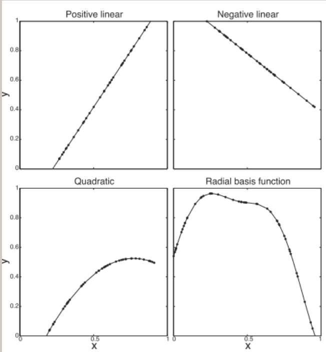
- Similar results in forecasting: trend-damping, importance of last point, added noise (Harvey et al.)
- Same in multiple-cue probability learning: linear mapping is the easiest to learn (Gluck et al.)
- Reinforcement learning: value function approximation has been proposed in ML, but not many investigations in human RL (Gershman and Daw)
- Dynamic control: linear function with only one dimension are easiest (Osman et al.)

Gaussian Process Models

Description

- Chris Lucas was the first one to propose this
- Non-parametric Bayesian model of function learning
- Adapts complexity to the data at hand
- Mixture of kernels with strong linearity bias

Overview

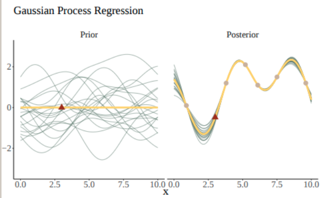


GP is a hybrid model

Description

- $y_* = \sum_{n=1}^{N} w_n k(s_n, s_*)$
- Each s_n is a previously observed input
- Weights are given by $\mathbf{w} = \left[\mathbf{K} + \sigma^2 \mathbf{I}\right]^{-1} \mathbf{y}$
- Mix of rule and kernel similarity

Overview



What's missing?

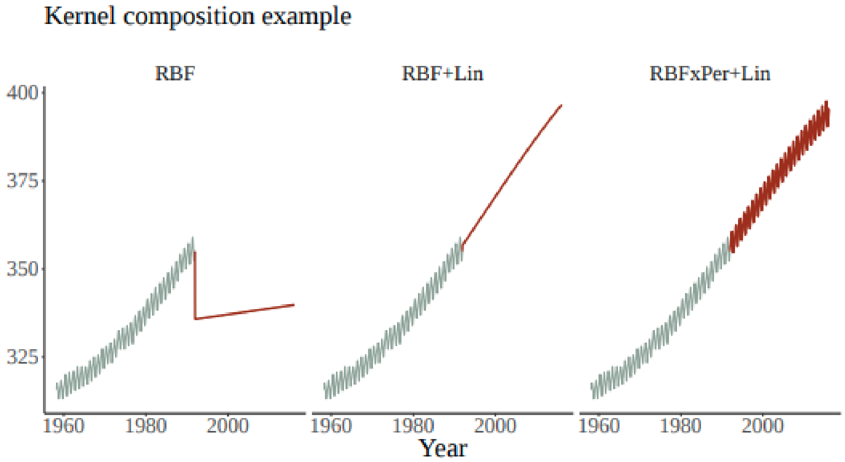
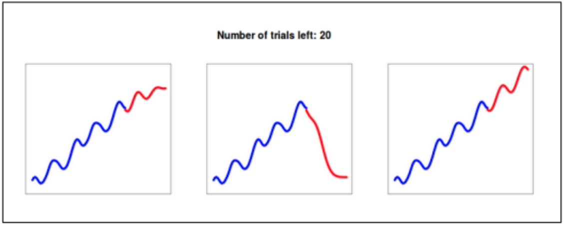
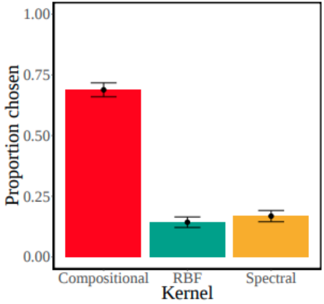


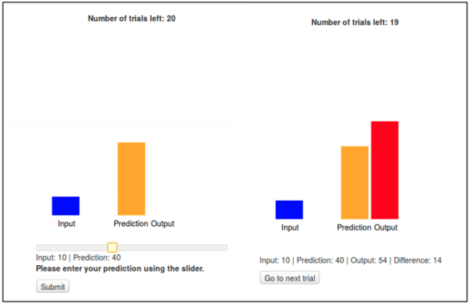
Figure: Compositional functions. Assume a grammar of Linear, Radial Basis Function and Periodic kernels. These can be added and multiplied together.

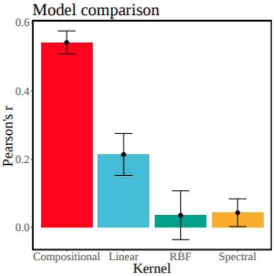
Simple choice task



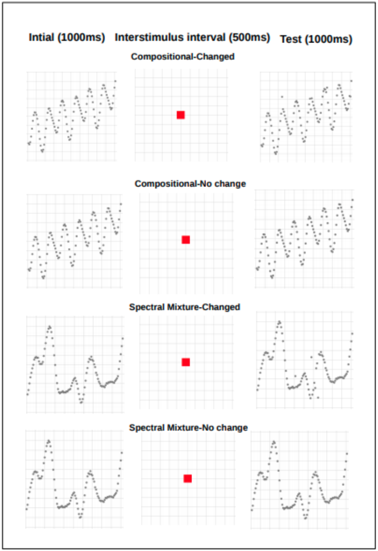


Classic task

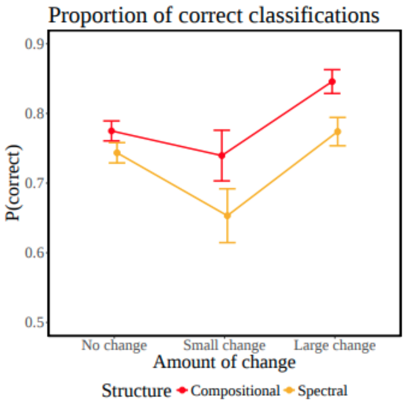




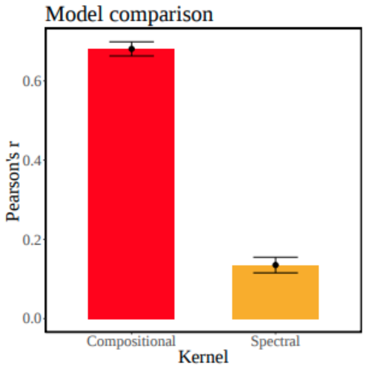
Memory task


Intial (1000ms)
Interstimulus interval (500ms)
Test (1000ms)
Compositional-Changed
Compositional-No change
Spectral Mixture-Changed
Spectral Mixture-No change

Memory task



Proportion of correct classifications
P(correct)
0.9
0.8
0.7
0.6
0.5
No change
Small change
Large change
Amount of change
Structure
Compositional
Spectral



Model comparison
Pearson's r
0.6
0.4
0.2
0.0
Compositional
Spectral
Kernel

Function learning in RL

Description

- Charley applied GP function learning to bandit tasks
- Put spatial correlation underneath options
- Nearby options produce similar rewards
- Participants can learn about this underlying function

Overview

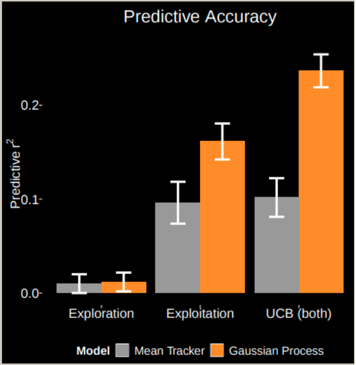
7	5	10	22	32	32	28	24	22	26	33
6	11	19	29	38	41	42	40	37	36	40
22	27	30	35	43	50	53	53	51	49	46
45	44	38	36	40	46	47	49	54	55	48
61	55	46	40	37	32	27	31	44	52	44
62	59	57	54	44	27	14	17	33	46	45
53	59	68	71	59	36	17	15	28	45	51
46	57	71	77	67	47	26	18	27	45	56
45	56	65	67	60	46	29	20	27	42	55
51	57	58	53	47	40	30	23	28	40	49
60	62	58	47	39	38	35	31	35	41	46

Function learning in RL

Description

- Gaussian Process model of function learning
- Combined with upper confidence bound sampling
- Solves generalization and exploration problem
- Works better than 26 alternative models

Overview

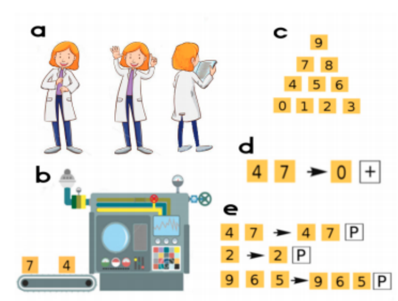


Problems with GPs

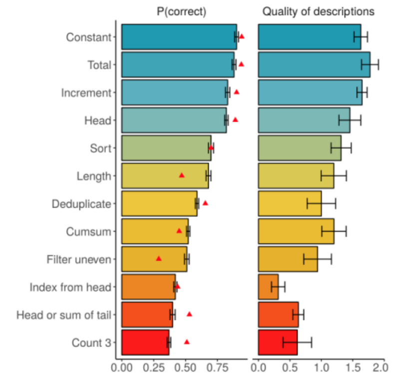
- Implausible scaling of complexity ($\rightarrow$ Susanne)
- Where do inductive biases come from? ($\rightarrow$ Shuchen + Akshay)
- Better to learn programs than functions? ($\rightarrow$ Alex)
- Better ways of exploration? ($\rightarrow$ Franziska + Lena)

Beyond simple functions: Programs

Overview

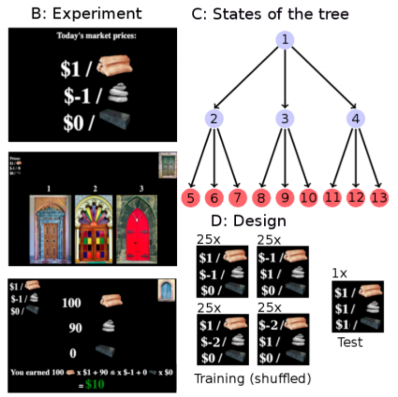


Results

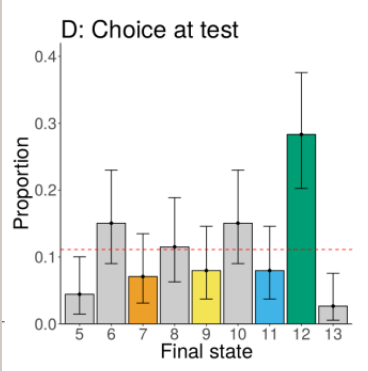


Beyond simple RL: Multiple tasks

Overview



Results



Conclusion

- Function learning is an important domain of cognitive psychology
- Several effects have been established
- Hybrid versions of rule-based and similarity-based learning work best
- Gaussian Process regression captures classic findings
- Can be extended to compositional inference and RL tasks
- Future directions: Learning programs and multi-task RL