Examining gradients of generalization in RL agents

Do RL agents follow a universal law of generalization?

Amanda Dsouza

adsouza41@gatech.edu

Psych 101 – Conditioned Reflexes

Image source: https://www.whole-dog-journal.com/behavior/dog-drooling-the-juicy-truth-about-why-dogs-slobber/

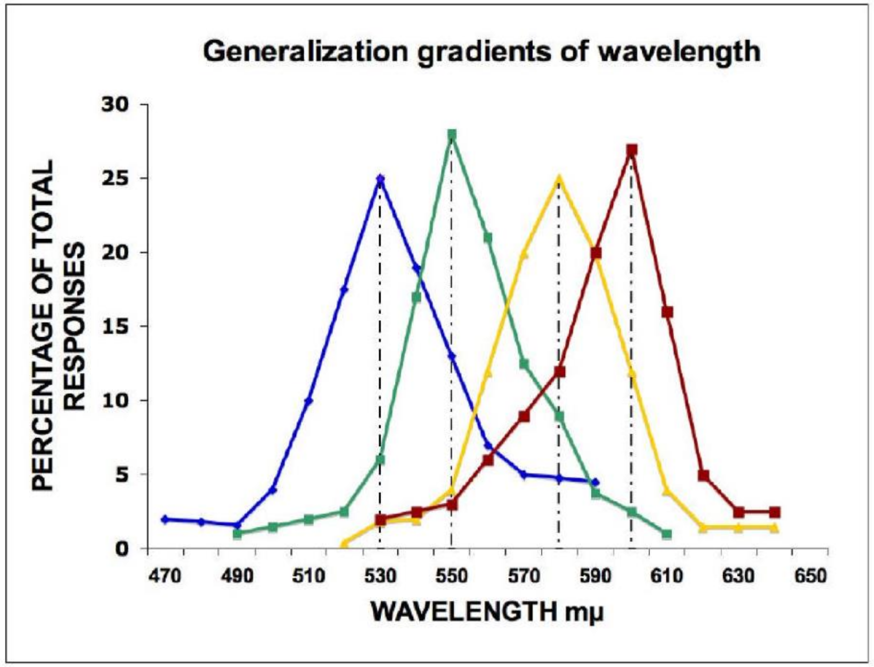
(1927) In Pavlov's classic conditioning experiments, a **neutral stimulus** (bell sound) associated with an unconditioned stimulus (food) was used to generate a conditioned response (salivating at bell sound).

In later experiments, conditioned responses were found to occur on test stimuli **different but similar** (e.g., pitch) to the original (training) stimulus.

Psych 101 – Gradients of Generalization

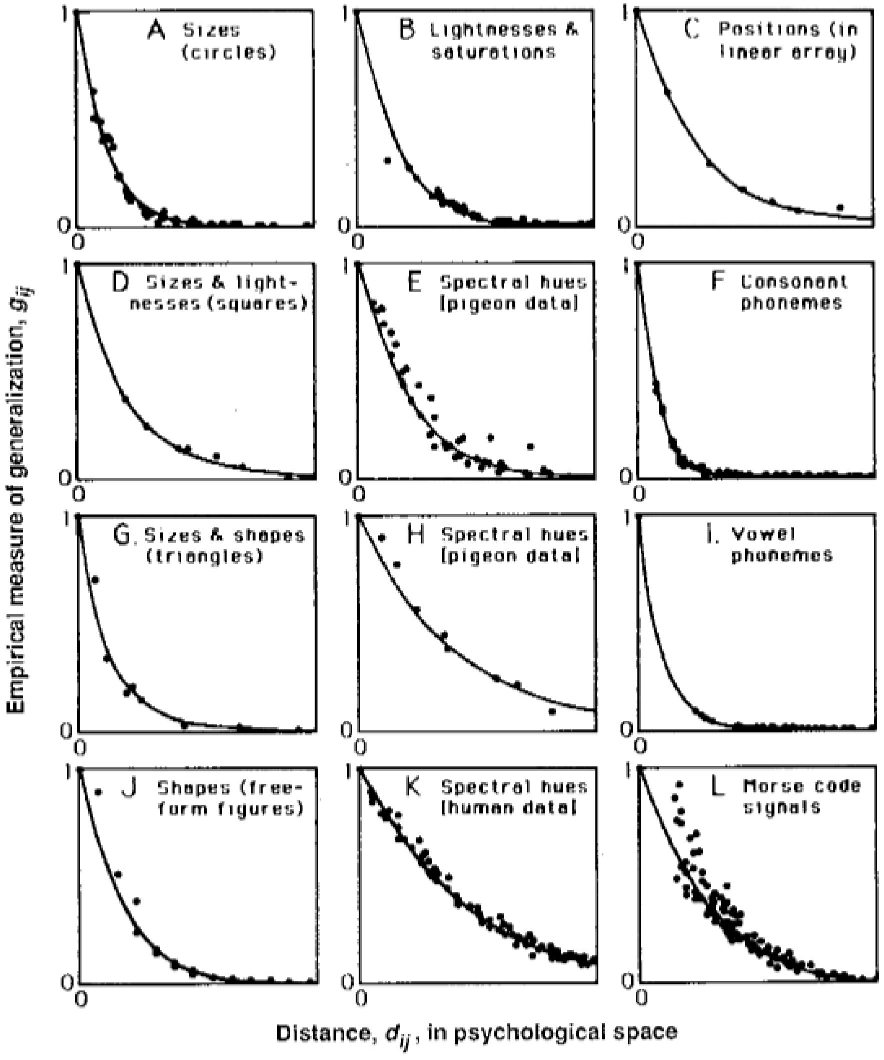
Image source: https://simple.wikipedia.org/wiki/Rock_pigeon

Followed by numerous experiments, analyzing "**gradients of stimulus generalization**", measuring the degree of learnt responses to distances between the test and original training stimulus



Guttman and Kalish 1956, Image source: Defining the stimulus—A memoir, H. Terrace

Psych 101 - Universal law of generalization



Toward a Universal Law of Generalization for Psychological Science, Roger Shepard

(1987) Following this work, Roger Shepard showed that there exists a **universal law of generalization**.

> "The probability to which a learnt response to a specific stimulus generalizes to another different stimulus depends on the "distance" between the stimuli and ***follows an exponential decay*** with this distance. Importantly, this distance measure is not in physical space, but one ***in psychological space.***"

Why should we care?

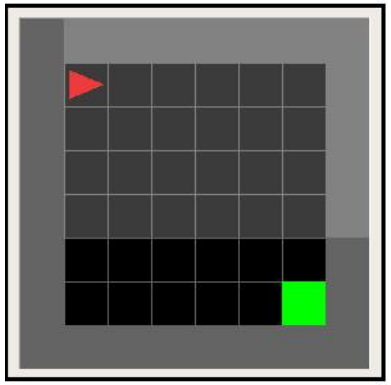
- Generalization in Psychology has a large body of research work, from which RL can draw insights. For e.g., studies of *peak shift* (heightened response to a stimulus not originally trained on, by introducing a negatively reinforcing stimulus). Could peak shift studies be used to encourage performance on stimuli, other than the one being trained on? (This could perhaps have applications in AI Safety.)

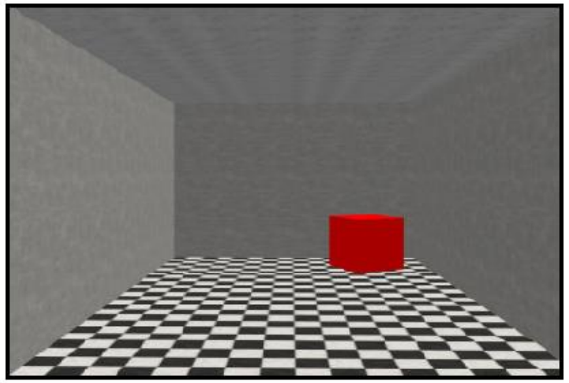
- Quantifying Generalization
  - It is often unclear how levels or environments in current generalization benchmarks differ from one another, beyond broad categories of easy/difficult, and what expected degrees of generalization should be.

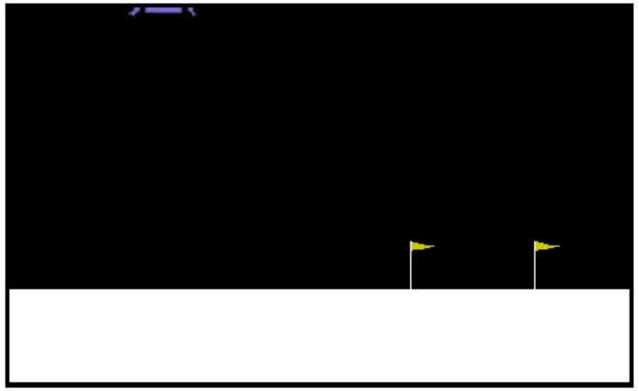
  - Can we use gradients of generalization to determine generalization guarantees, similar to scaling laws?

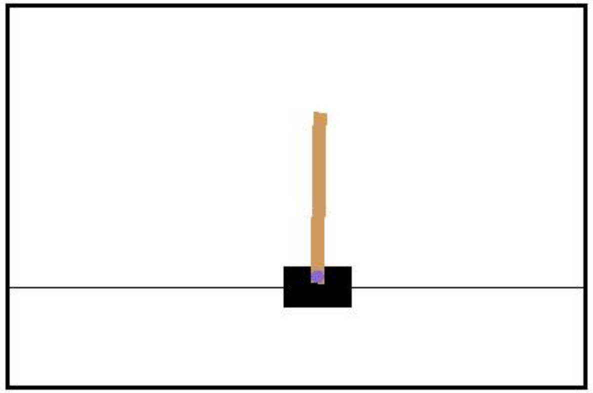
- Power laws (e.g., scaling, inverse scaling laws) are all the rage right now?

Examining gradients of generalization in RL agents





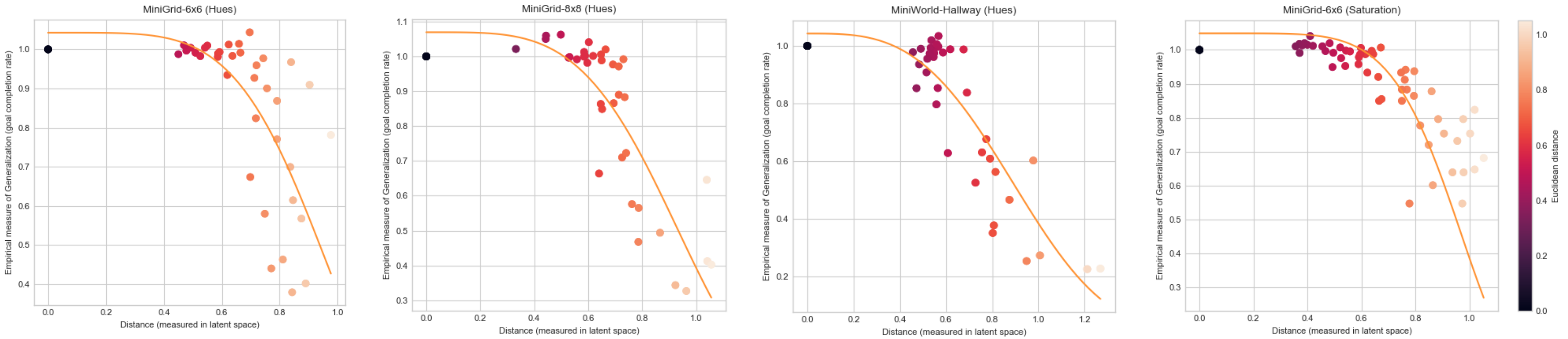




Shepard’s experiments were conducted on humans and animals on a wide range of stimuli such as colors attributes (lightness, saturation), sizes, spectral hues, phonemes, shapes, and Morse code signals.

Experiments are conducted by modifying three simple environments, MiniGrid and MiniWorld (Minimalistic 3D Environment with egocentric views) and OpenAI Gym Classic Control (Lunar Lander, CartPole).

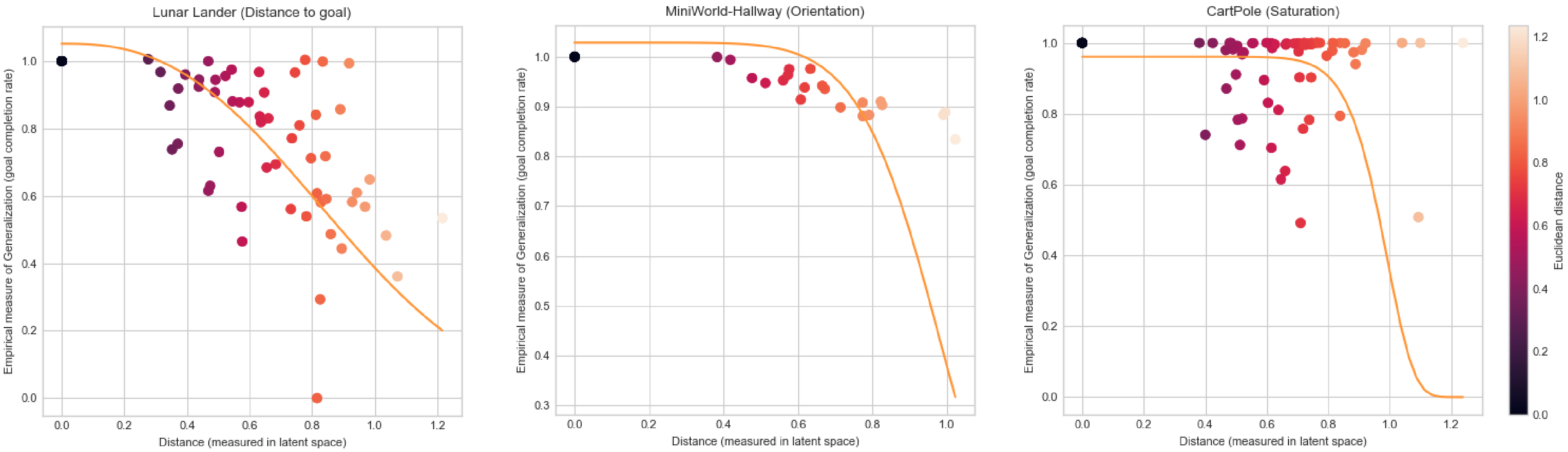
Examining gradients of generalization in RL agents



Results averaged over 5 seeds, each with mean reward over 100 episodes

Experiments varying neutral and symmetric stimuli, such as hue and saturation values of goal (tile, box) in MiniGrid and MiniWorld environments exhibit stretched exponential decays with increasing distance in latent space. (A stretched exponential has the form $f_\beta(t) = e^{-t^\beta}$).

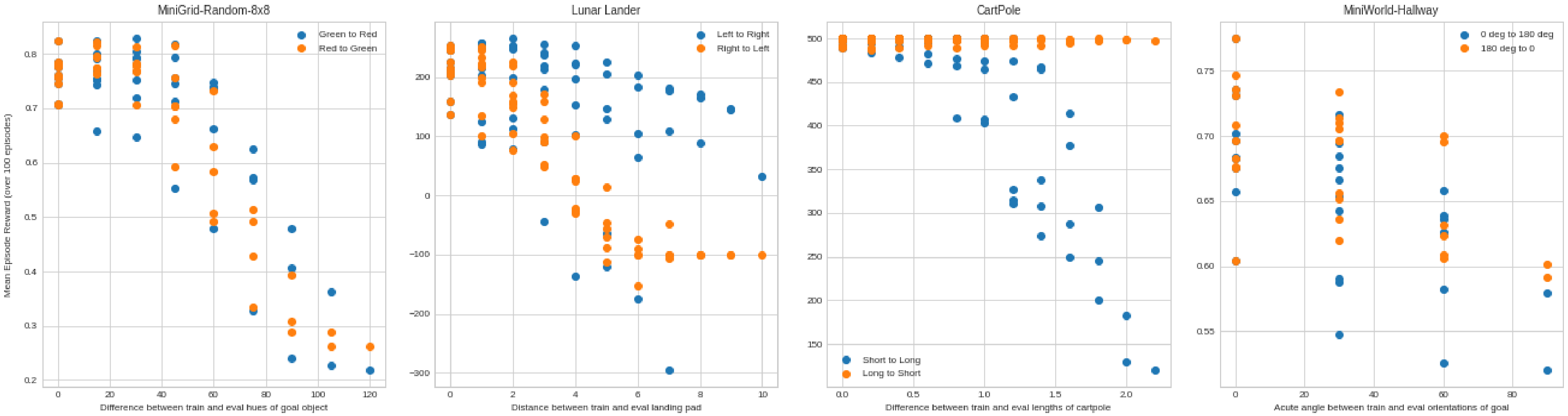
Examining gradients of generalization - II



On stimuli that are not neutral or symmetric, but instead, change the task (in complexity or otherwise), we cannot recover an exponential curve. Given that some orientations (less occluded) ought to be more easily learnable than others, a stretched exponential may not be recovered. It is not possible to recover an exponential generalization curve on cartpole lengths due to the asymmetric nature of the stimulus, i.e., it is easier to generalize from longer to shorter poles than vice-versa. Read more here.

Examining gradients of generalization - III

Plot of true distance to degree of generalization. (a) Reference stretched exponential decay of generalization in hues (b) Distance of lunar lander to landing pad. Surprisingly, it is easier to generalize from Left -> Right, than Right -> Left. This could be due to the dynamics of the lander. (c) Lengths of CartPole. Longer cartpoles can generalize to shorter ones. (d) Orientation of box (goal) in MiniWorld-Hallway, linearly decaying in acute angles between train/test.



How to uncover the universal law

- For normalized generalization data G, on a set of stimuli S (s+ as the original trained stimuli), and a distance metric *d*, we want to recover a function *f* as,

$$d(s+, s-) = f^{-1}(g_{s+,s-}) \qquad g \in G, d : S \times S \rightarrow R$$

- This can be done by using non-metric multidimensional scaling (NMDS), which preserves the ordering of the similarity in data.

- On an *n* X *n* symmetric matrix of (normalized) generalization measures *g_ij*, NMDS finds a lower dimensional space in some *k* (*k* << *n)* dimensions. Points in this space can now represent distances between the original points and are invariant to underlying experiment data.

- Then, using some metric distance (Shepard shows that Euclidean and Manhattan distances work well for stimuli data), one can uncover the universal law.

Findings

- Main takeaway is that there may be some fundamental similarities between how biological and artificial agents learn, that allow for similar generalization patterns.

- Experiments with stimuli that are independent of the task (**neutral stimuli**) exhibit a similar (stretched) exponential decay as the universal law in behavioral experiments on humans/animals.

- When stimuli are related to the task, altering the learning capacity or complexity of the task, the gradients exhibit different curves.

- Only neutral and/or symmetric stimuli (such as the different pitch sounds in Pavlov's experiment) can produce such behavior.

- What constitutes neural/symmetric stimuli is an important distinction between the gradient experiments on humans/animals, and on artificial agents. Stimuli such as size and orientation were found to exhibit the same universal law, when performed on human/animals. However, in RL agents, these stimuli do not show the same curves, since they alter the complexity of the task in one direction. (e.g., more complex environments typically generalize better to simpler ones).

Thoughts

- More details in blog

- How this could be improved?
  - More experiments with different stimuli?
  - Statistical tests of fitting to distribution
  - Exploring some of the critiques of the universal law – metrics used/proposed, properties of metrics used.

- What are important directions for research related to this work?
  - Related generalization experiments – peak shifts etc.

References

- Toward a Universal Law of Generalization for Psychological Science, Roger N. Shepard, Science, New Series, Vol. 237, No. 4820. (Sep. 11, 1987), pp. 1317-1323.
- PSY402 Theories of Learning, https://www.cpp.edu/~nalvarado/PSY402%20PPTs/New%20Klein/PDFs/KleinCh10.pdf
- Decision-Making and Learning: The Peak Shift Behavioral Response, S. K. Lynn, Boston College, 2010 Elsevier Ltd

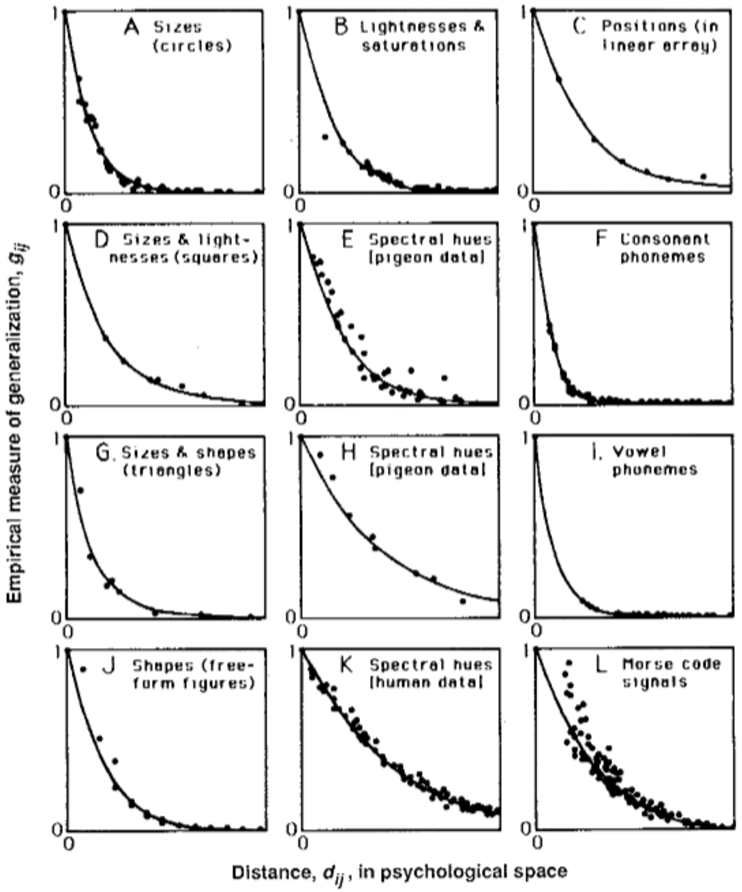
Appendix - Universal law of generalization

- Previous attempts at measuring "gradients of generalization" used physical measures of differences between stimuli (frequency / size / wavelengths etc.).

- Even though physical differences lead to an overall decrease of generalization with increasing distance, the decrease is not necessarily monotonic nor invariant.

- Shepard, therefore, sought to find a monotonic and invariant function, whose inverse transformed the observed generalization data into distances in some space (he termed as psychological space). [Can be thought of as latent space in ML terms]

- Further, he found that the exponential decrease in this "psychological space" follows universally among different stimuli, sensory modalities, and across multiple species.



Toward a Universal Law of Generalization for Psychological Science, Roger Shepard