# Examining gradients of generalization in RL agents

Do RL agents follow a universal law of generalization?

Amanda Dsouza

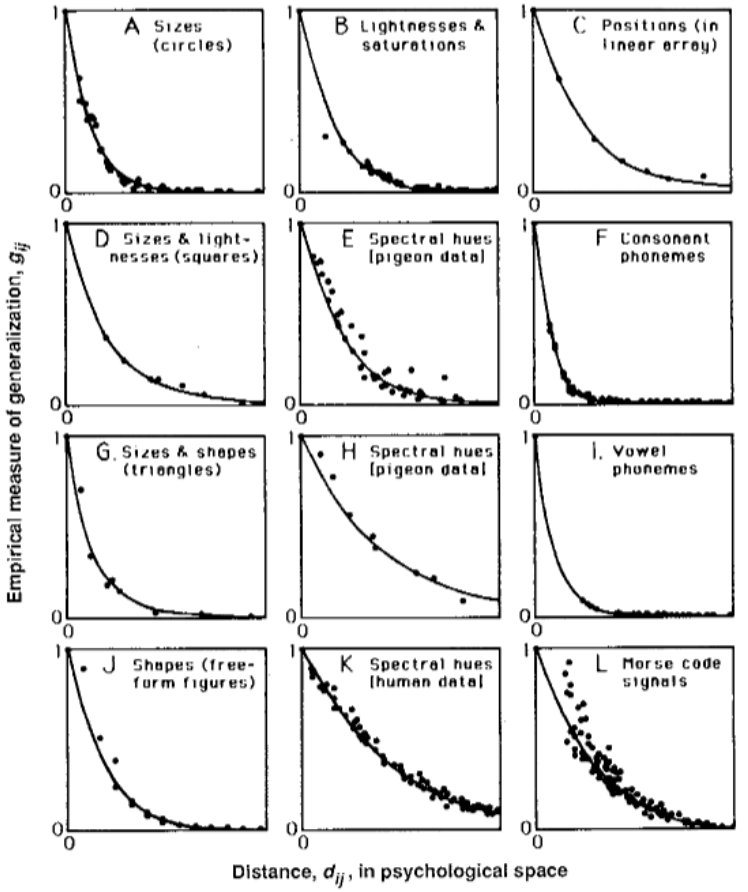
adsouza41@gatech.edu

# Background - Conditioned Reflexes

- (1927) In Pavlov's classic conditioning experiments, a **neutral stimulus** (bell sound) associated with the unconditioned stimulus (food) was used to generate a conditioned response (salivating at bell sound).
- In subsequent experiments, conditioned responses were found to occur on test stimuli different but similar (in pitch, for instance) to the original conditioned (training) stimulus.
- This led to numerous experiments, analyzing "**gradients of stimulus generalization**", measuring the degree of learnt responses to distances between the test and original training stimulus.
- (1987) Following this work, Roger Shepard showed that there exists a **universal law of generalization**.
- According to the law, 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*.

# Background - 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

# Why should we care?

- 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?
- Generalization in Psychology has a large body of research work, from which RL can draw insights. For instance, peak shift has been studied extensively (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.)
- Power laws (e.g., scaling, inverse scaling laws) are all the rage right now?

# How to uncover the universal law

- For 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,

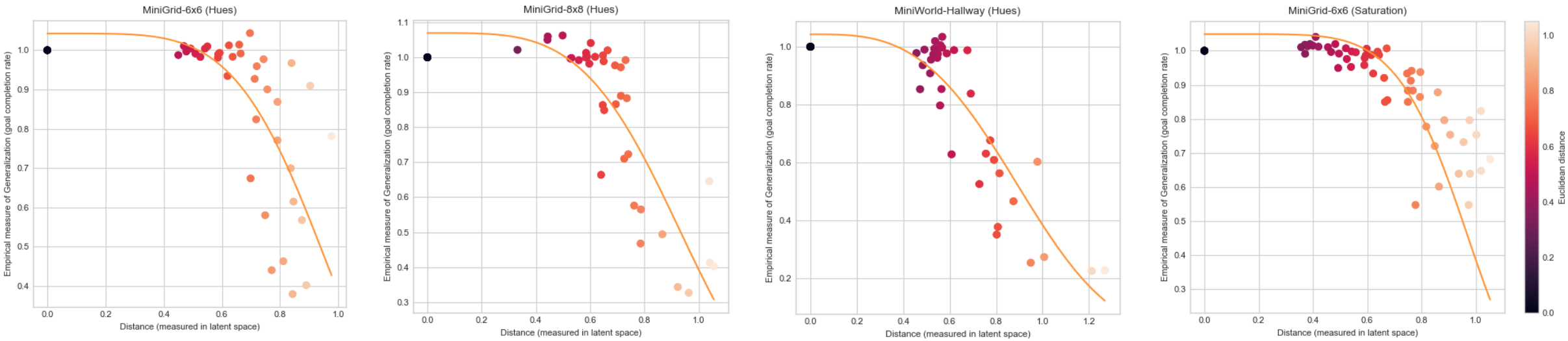
$$d(s+, s-) = f^{-1}(g_{s+,s-}) \qquad g \in G, d : S \times S \to R$$

- 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.

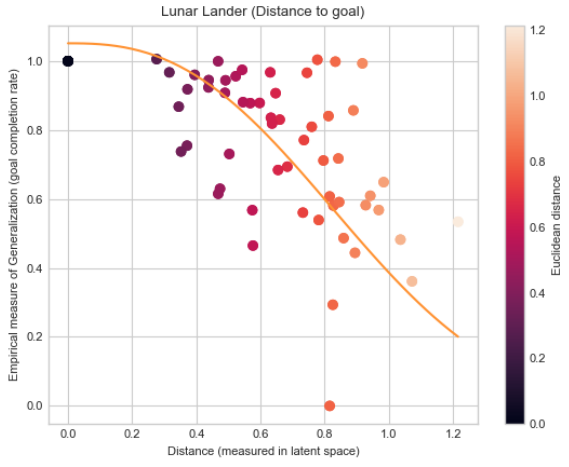
# Examining gradients of generalization - I




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

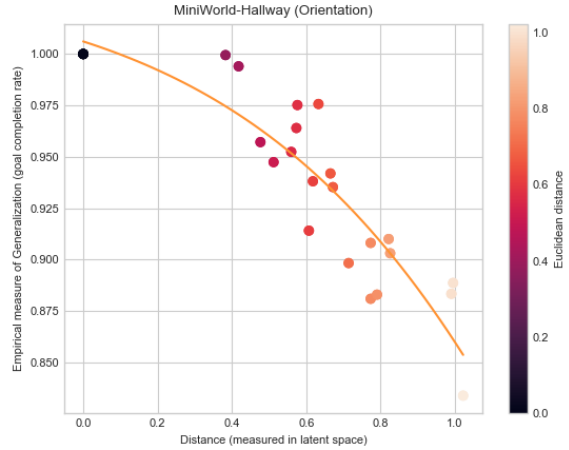
Experiments on varying hue and saturation values of goal (tile, box) in MiniGrid and MiniWorld (Minimalistic 3D Environment with egocentric view observations and discrete actions) environments exhibit stretched exponential decays with increasing distance in latent space. (A stretched exponential has the form $f_\beta(t) = e^{-t^\beta}$ ).

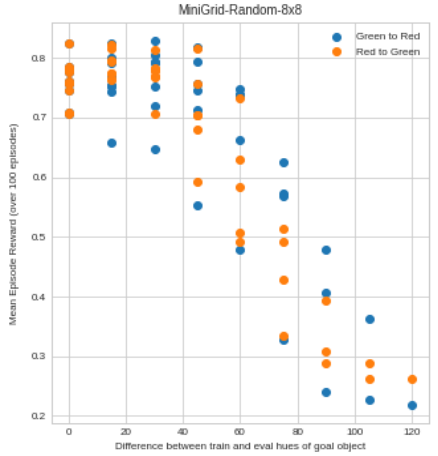
# Examining gradients of generalization - II

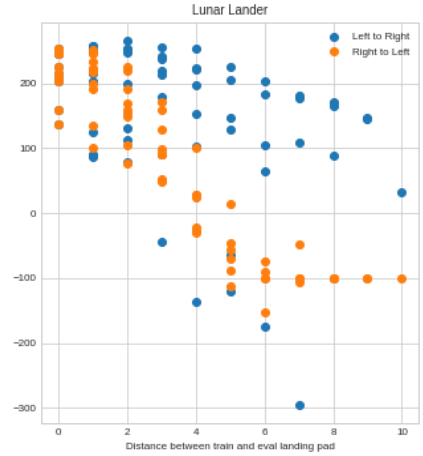


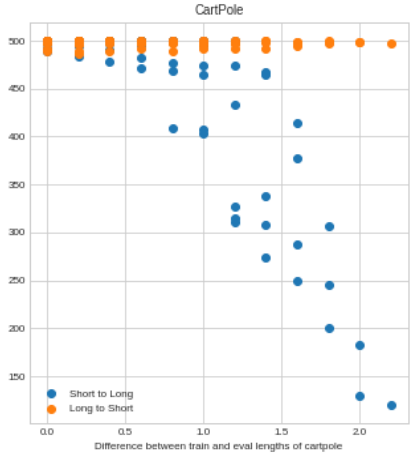

LEFT: Distance of lunar lander from landing pad, and orientation of box (goal) in MiniWorld-Hallway do not strictly follow a stretched exponential decay. A logistic curve is fitted on orientation data.

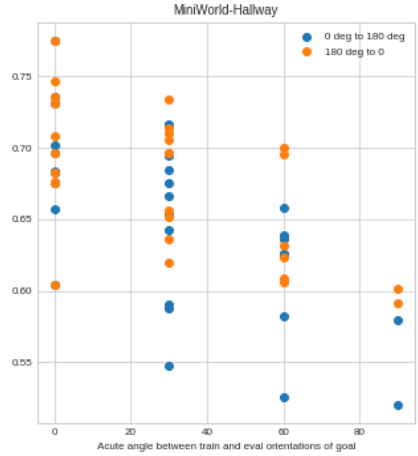
BOTTOM: Plot of true distance to degree of generalization. (a) Reference stretched exponential decay of generalization in hues (b) Distance of lunar lander to landing pad. Surprisingly, it is easier to generalize from Left -> Right, than Right -> Left. This could be due to the dynamics of the lander. (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.



# Findings

- On experiments with stimuli that are independent of the task (**neutral stimuli**), we find they 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 stimuli (such as the different pitch sounds in Pavlov's experiment) can produce such behavior.

- What constitutes neural 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).

# 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