[Supplementary Information]

Human Category Learning in Reversal Tasks: Dynamic Learning Rates Help Overcome Catastrophic Interference in Connectionist Networks

Santiago Castiello1\*, Andrew R. Delamater2

1Deparment of Psychiatry, Yale University, US

2Deparment of Psychology, Brooklyn College and Graduate Center of CUNY, US

\*Corresponding author: santiago.castiellodeobeso@yale.edu

# Connectionist Model

The model was implemented with code written in R and is publicly available at the following website: <https://github.com/santiagocdo/ALANN> (check folder *training\_files*: *category\_total\_reversal.csv* and *category\_partial\_reversal.csv*). The connectionist model consists of a three-layer feedforward neural network, where activations go bottom-up (from CS to US representations) and learning top-down. The layers are input (L = 0), hidden (L = 1) and output (L = 2). This model can be described with two rules: the activation rule (that calculates unit activations) and the learning rule (that determines changes in weights or connections between units).

The connectivity constraint between the input and the hidden layer given the visual, multimodal, and auditory pathways (Delamater, 2012) is encoded in the **C** matrix (size: number of input units x number of hidden units), where 1 represents an existent connection and 0 a non-existent connection; hence, connections between visual input units to auditory hidden layer pathway units will be equal to 0, and similarly for auditory inputs and visual pathways.

The model computes weight matrices for layers 1 and 2, i.e., between input and hidden (L = 1), and between hidden and output (L = 2), at every trial *t.* The values of at *t* = 0 are random variables from a uniform distribution between -0.5 and 0.5. To maintain the partial connectivity between input and hidden  is multiplied (via Hadamard product or elementwise multiplication) with **C**. On the one hand, maps the connections’ weights between input and hidden units, thereby, the matrix size is the number of inputs x number of hidden units. On the other hand, represents the weights between hidden and output units, hence, the matrix size is the number of hiddenx number of outputs. At every *t*, the CS information is encoded in the activation of the input layer, , where this vector contains 0s and 1s, representing presences and absence of CSs). The USs are encoded in a vector (size is equal to the number of USs)[[1]](#footnote-1).

*Activation Rule*. The first rule of this model specifies how units are activated. The activation of the hidden and output layers is given by a sigmoidal function that receives the net input (**n**) vector for the current layer L, i.e., :

(1).

The activations are a real number between 0 and 1. We subtract 2.2. from to shift the sigmoidal curve to the right and obtain low activation levels when elements is 0. The net input for the hidden layer is:

(2),

where is a matrix multiplication between a transpose () column vector (i.e., a row vector), and the matrix.

*Learning rule*. The second rule of the model specifies how changes at every *t*. This rule is an adaptation of backpropagation (Rumelhart et al., 1986). In general terms, the weights are the sum of the current weights and the change of those weights at *t*. Hence:

(3),

Where is the Hadamard product or elementwise multiplication, and the partial connectivity only applies to input layer (L = 0) to hidden layer (L = 1) connections within the weight matrix . This ensures that weights between units with impermissible connections (e.g., auditory inputs to visual pathway hidden units) are fixed at 0. Changes in weights are determined by:

(4),

where and are free parameters, representing the learning rate (assumed to be 0.3) and the momentum decay (assumed to be 0.9), respectively. Finally, is a vector of delta values, and encodes the prediction errors that are back-propagated throughout the network, following the next equation:

(5),

where is a vector which encodes with 0s (absent) and 1s (present) the USs at every *t*.

# Instructions

## Experiment 1a and 1b

In this experiment, you will be presented with a series of abstract images that represent the molecular structure of various natural objects. Your task will be to learn from which of two regions in the world these objects come (Northern Hemisphere or Southern Hemisphere).

More specifically, you will see 1 of 8 different abstract images at a time and be asked to indicate whether you think that image reflects an object taken from the Northern or Southern Hemisphere. Choose the Left Arrow Key for Northern or the Right Arrow Key for Southern Hemisphere. At first, you will need to guess, but you will be provided with feedback after your answer to help you learn which objects come from Northern or Southern Hemispheres.

Your response times are also important. Please make your response choices as quickly, but also as accurately, as you can. Your feedback will display the time (in sec) that it took for you to reach your decision, and also if your choice was correct (with a high pitch sound) or not (low pitch sound).

There will be a break halfway through.

Press the space bar when you are ready to begin.

## Experiment 2

[categorySimilarity]

In this experiment, you will be shown pairs of abstract images. Your task is to rate how similar or different the two images appear to you.

The experiment consists of two parts. In the first part, you will see and rate 56 pairs of images, and in the second part, another 56 pairs. Below each pair, there will be a slider ranging from "very different" to "very similar." Use your mouse to select a position on the slider that best represents how similar or different you find the two images are relative to one another.

There will be a break between both parts.

Press the space bar when you are ready to begin.

[fractal\_similarity]

In this experiment, you will be shown pairs of abstract images. Your task is to rate how similar or different each pair appears to you.

The experiment consists of two parts. In the first part, you will see 56 pairs of images, and in the second part, another 56 pairs, individual images may repeat between pairs. Below each pair, there will be a slider ranging from "very different" to "very similar." Use your mouse to select a position on the slider that best represents how similar or dissimilar you find the images.

There will be a break between both parts.

Press the space bar when you are ready to begin.

# Informed Consent

This research investigates the psychological processes used when people learn to identify objects in the world. You are being asked to participate in this research study because you are a normal healthy adult, and we wish to better understand basic learning processes in your population. The purpose of this research is to gain more knowledge about the cognitive processes involved in simple forms of associative learning.

If you agree to participate, we will ask you to perform in a simple computer task that will last approximately 12 minutes. In this task, you will see a series of abstract images presented individually on the screen and your task will be to learn to choose one of two response options for each image. Also, you will be asked to respond quickly and accurately on your computer keyboard when the image appears.

• Risks/Discomforts: There are no risks for participating in this study beyond those associated with normal computer use over a 15 min period.

• Benefits: This research is not designed to directly benefit you, but your help with this study will advance basic science on the cognitive processes involved in predictive learning in normal healthy individuals. Ultimately, this research could lead to a better understanding of some of the associative learning processes that are negatively impacted by various psychological conditions (such as aging, dementia, etc).

• Confidentiality: This study does not collect identifying information, and all data collected will remain anonymous. We will ask about your gender, age, and nationality, and we will record your performance in the task itself. However, this information will not be linked directly to any individual participant. The data we obtain will be stored indefinitely and may be shared publicly via online repositories for findings that are ultimately published.

Your participation in this research is completely voluntary, and you will be able to stop at any time without penalty. If you have any questions, you can contact: Andrew R. Delamater (andrewd@brooklyn.cuny.edu). If you have any questions about your rights as a research participant or if you would like to talk to someone other than the researcher, you can contact CUNY Research Compliance Administrator at 646-664-8918 or HRPP@cuny.edu.

If you wish to participate in the study, please press the spacebar for additional instructions.

# Analysis

## Experiment 1a: Sensitivity Analysis of the Reaction Time

*RT transformed with Natural Logarithm*. We ran the same LMM with condition (total versus partial), Blocks, and its interaction. We found no main effects in Total reversed versus Partial reversed [] nor Blocks [], and no Interaction [].

*RT transformed with Root Square*. We ran the same LMM with condition (total versus partial), Blocks, and its interaction. We found no main effects in Total-Condition [] nor Blocks [], and no Interaction [].

## Experiment 1b: Sensitivity Analysis of the Reaction Time

*RT transformed with Natural Logarithm*. We ran the same LMM with condition (total versus partial), Blocks, and its interaction. We found the main significant effects in Total reversed versus Partial reversed [] and Blocks [], but no Interaction [].

*RT transformed with Root Square*. We ran the same LMM with condition (total versus partial), Blocks, and its interaction. We found a main effect in Total reversed versus Partial reversed [], but not in Blocks [] nor the Interaction [].

## Experiment 2: Multidimensional Scaling (*k*=3)

**A diagram of a distance

AI-generated content may be incorrect.**

**Figure S1**. *Category Similarity Analysis, using two psychological dimensions (multidimensional scaling with k=3) in the psychological space.* *Left panels (****A*** *and* ***C****) represent Experiment 1a results for the exemplars for Set 1, and the right panels (****B*** *and* ***D****) represent the comparisons conducted with exemplars in Set 2. Each dark point represents a comparison for each participant, i.e., for each of the 28 stimulus comparisons. Boxplots represent the median, interquartile (IQR) range, and the whiskers are 1.5 the IQR. \*\*\*: p < .001, \*\*: p < .01, \*: p < .05.*

1. All vectors are considered to be column vectors. Hence to get a row vector we transpose them, e.g. **x** (column vector), **x**T (row vector). [↑](#footnote-ref-1)