

# Investigating the Diversity of Evolved Cognitive Models in Delayed Match to Sample Experiments

Peter C.R. Lane<sup>1\*</sup>, Laura K. Bartlett<sup>2</sup>, and Fernand Gobet<sup>2</sup>

<sup>1</sup>*Department of Computer Science, University of Hertfordshire*

<sup>2</sup>*Centre for Philosophy of Natural and Social Science, London School of Economics*

\*corresponding author: p.c.lane@herts.ac.uk

The construction of programs to model cognitive behaviour can be semi-automated, using program synthesis techniques such as genetic programming. However, these techniques can produce many candidate solutions of good fitness making it difficult to find the important similarities and differences which can explain model behaviour and lead to theoretical insights. Here we show how the diversity of generated models can be assessed using the Silhouette Index measure of cluster quality; the quantitative measure helps guide theorists to important qualitative differences in the models. Such differences can help inspire novel cognitive theories.

Keywords: cognitive models, explainable AI, genetic programming, silhouette index, visualisation

## Introduction

Computer-based tools can assist scientists in the development of new theories. Such tools can analyse large datasets and reveal candidate solutions to problems beyond the reach of manual search. In this abstract, we use our model-discovery system, Genetically Evolving Models in Science (GEMS) [1,2], to develop cognitive models of two related but subtly distinct experiments. These experiments are examples of the Delayed Match to Sample task (DMTS), a popular experimental format for studying short-term memory and decision-making abilities. We create models from two separate experiments using the DMTS format: experiment 1 is by Chao et al. [3] and experiment 2 by Edwards et al. [4]. GEMS has been used to develop models for both sets of results previously [1, 5], and here we build on these results to provide a more detailed analysis and comparison.

## GEMS System

The DMTS task has a format consisting of the presentation of a target image, a delay where no image is visible, followed by the presentation of two stimulus images: the participant must indicate which of the two stimuli is identical to the original target as quickly as possible. The two examples of this experiment used different kinds of images and also different lengths of delay: Chao et al. [3] showed images of animals, tools and faces, with a delay of 0.5 seconds, whereas Edwards et al. [4] showed images of abstract shapes with a delay of 5 seconds. The aim of the models is to simulate human behaviour both in accuracy and in response time: for experiment 1 average accuracy was 95.7%, response time 767ms, and for experiment 2 average accuracy was 94%, response time 485ms.

In order to model these experiments, we construct control programs for a simple cognitive architecture. The architecture has input/output operators, a short-term memory (STM) and an attended item. All operators require a particular period of time to function, and this time is taken from earlier research in psychology [6], e.g. the time to recognise and input a stimulus is 100ms, and the time to compare items in STM is 70ms. Some operators permit the model to simply wait for a period of time. Combinations of these operators form control programs which potentially simulate the behaviour of a participant in one of the two experiments. Due to the timing of operators, we can both predict the accuracy and the simulated reaction time of each model, and so compare these measurements to those of humans.

The similarity between two programs is computed by extracting all program nodes and their immediate children, and then computing the Jaccard Index of the two sets (size of intersection / size of union); distance is

then 1-similarity [1]. For example the following program is converted into eight segments of two parts and six individual node names:

```
(if (compare-1-2)
  (prog2 (input-right) (input-left))
  (input-target))
parts: (if compare-1-2 prog2 input-target) (prog2 input-right input-left)
names: if compare-1-2 prog2 input-target input-right input-left
```

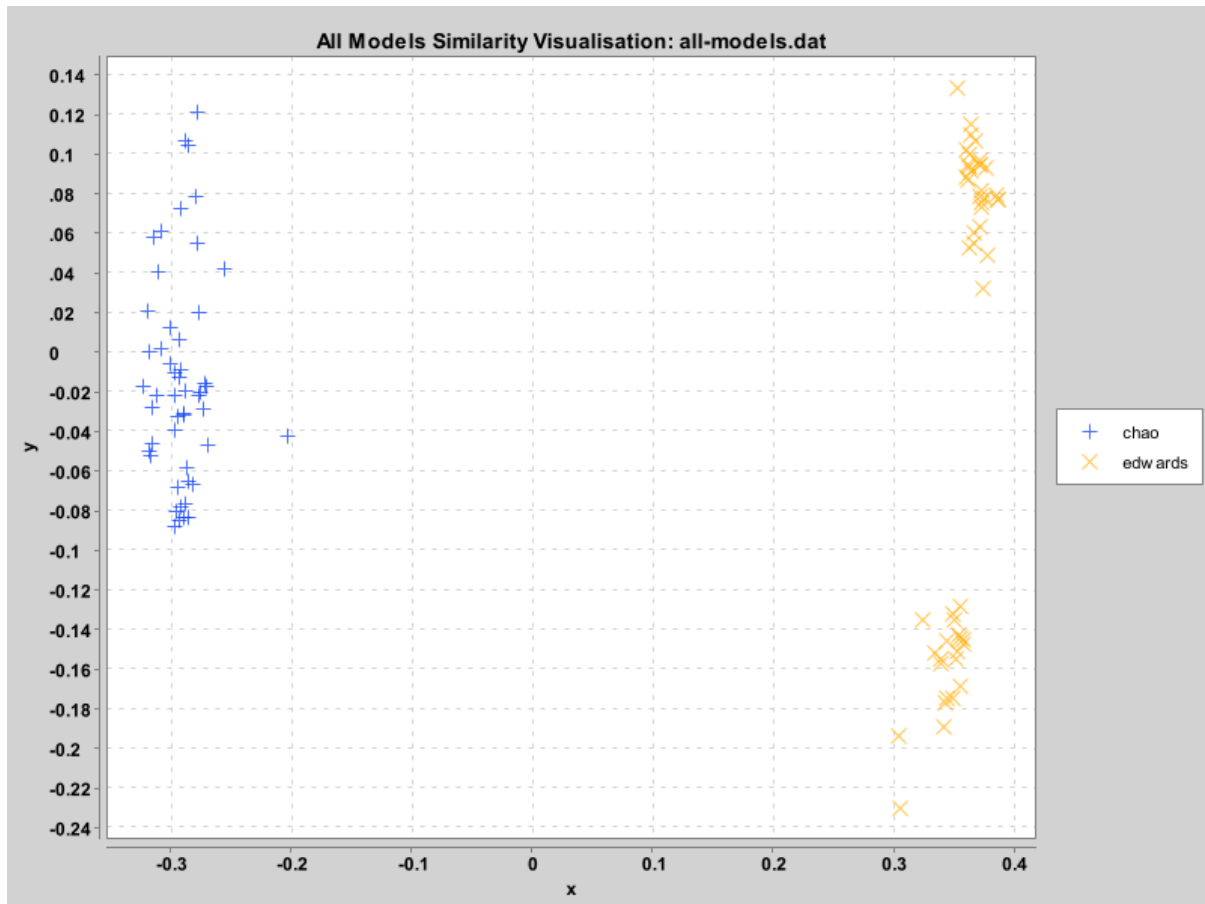


Fig. 1. Visualisation of model similarity for models evolved to fit two different experiments. Models on the left are a fit for experiment 1 and models on the right are a fit for experiment 2.

### Simulation Results

Genetic Programming is used to generate candidate models for each of the two experiments, using a combined fitness function computed from the difference of each model's accuracy and response time to those observed in the experiments with human participants. For each experiment, a population of 5000 individuals was evolved for 250 generations. The final population was then filtered, to extract all the "good" models - a good model is defined as one with an overall fitness of less than 0.1. Post-processing [1,2] then removes dead code and replaces time-only code with wait operators to remove duplicate models; the number of candidate models is reduced from 6,500 models in total to a more manageable, but still large number of distinct models - 583 models for experiment 1 and 177 for experiment 2.

## Results and discussion

Fig. 1 shows the best 50 models generated for each of the two experiments: their mutual similarity is converted into a 2D coordinate using multidimensional scaling. As is evident visually, models from the two experiments are distinct from each other: those modelling the first experiment, from Chao et al. [3], falling on the left of the figure, and those modelling the second experiment, from Edwards et al. [4], falling on the right. It is also apparent that the models from the second experiment fall into two groups, based on the vertical separation.

Although visual examination reveals the structure of the solution space in this case, a quantitative technique is preferred to automate the recovery of different kinds of models. In this case, we use k-means clustering using the Silhouette Index as a measure of cluster quality to determine that 3 clusters is optimal. The scientist is then presented with a typical example of each of the 3 clusters for qualitative analysis.

## Conclusion

We have compared models developed for two different but related experiments. These models have evolved similar but distinct solutions to the task. Analysis techniques have identified strong similarities where expected, in the initial visual scanning parts of the experiments, and some differences in the later processing. Although relatively simple as models, the analysis techniques are important in showing how the large number of candidate models produced by a program synthesis tool can be automatically processed and converted into explainable results, highlighting qualitatively distinct forms of behaviour. Cognitive scientists are thus freed from the time-consuming and error-prone tasks of constructing multiple candidate models, and instead are able to spend more time considering the implications of the analysis for the development of novel theories of human behaviour.

More information about the GEMS project and the software underlying the experiments in this paper can be found at: <https://gems.codeberg.page/>

**Acknowledgements** This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. ERC-ADG-35002).

## References

- [1] P.C.R. Lane, L.K. Bartlett, N. Javed, A. Pirrone and F. Gobet, 'Evolving understandable cognitive models', in T.C. Steward (Ed.), *Proceedings of the 20th International Conference on Cognitive Modelling*, pp.176-82, 2022.
- [2] P.C.R. Lane, F. Gobet, A. Pirrone, L.K. Bartlett and N. Javed, 'Synthesising cognitive models with genetic programming', *Proceedings of Abstracts, School of Physics, Engineering and Computer Science Research Conference*, 2023.
- [3] L. Chao, J. Haxby, and A. Martin, 'Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects', *Nature Neuroscience*, 2:913–20, 1999. <https://doi.org/10.1038/13217>
- [4] J.E. Edwards, T.W. Boyer, Z.K. Bell and B.R. Sturz, 'Isolated processing of geometric shapes and their corresponding shape words: Evidence from a delayed match-to-sample task', *Journal of Experimental Psychology: Human Perception and Performance*, 42(8), 1088–1103, 2016. <https://doi.org/10.1037/xhp0000210>
- [5] L.K. Bartlett, A. Pirrone, N. Javed, P.C.R. Lane and F. Gobet, 'Genetic programming for developing simple cognitive models', in M. Goldwater, F.K. Anggoro, B.K. Hayes, & D.C. Ong (Eds.), *Proceedings of the 45th Annual Meeting of the Cognitive Science Society*, pp.2833-2839, 2023. <https://escholarship.org/uc/item/08x8m02w>
- [6] S.K. Card, T.P. Moran and A. Newell, *The Psychology of Human-Computer Interaction*, 1983. (Hillsdale, N.J.: Erlbaum)