The Effect of Representational Compression on Flexibility Across Learning in Humans and Artificial Neural Networks

Mia Whitefield
University of Oxford, Oxford, United Kingdom

Christopher Summerfield
Oxford University, Oxford, United Kingdom

Abstract

Humans can generalise from past experiences to novel situations as well as revise prior knowledge to flexibly adapt to changing contexts and goals. The representational geometry framework formalises how information is structured in the brain and suggests that abstraction involves a trade-off between generalisation and flexibility. However, how task representations evolve across learning and relate to behaviour remains unclear. Here, we tested the hypothesis that representational compression of task representations across learning underlies this flexibility impairment. Using an extra-dimensional shifting task, we manipulated the pretraining length to control the degree of compression. In both humans and artificial neural networks, longer pretraining was associated with decreased flexibility. Network dynamics indicated that greater compression incurred a higher representational reorganisation cost, limiting flexibility. Introducing an auxiliary reconstruction loss maintained higher dimensionality and mitigated the flexibility impairment. These findings suggest representational compression constrains flexibility, and preserving representational richness enhances flexibility.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*