Abstract

Cognitive systems, whether biological or artificial, operate under inherent resource limitations that necessitate efficient coding strategies. This thesis explores the fundamental principles governing efficient coding through the framework of Rate-Distortion Theory (RDT), investigating how limitations on information representation shape the geometry of latent spaces in generative models and contribute to the emergence of cognitive abilities, specifically number sense. The central argument is that resource constraints, formalized by RDT, impose a trade-off between the fidelity of information representation and the capacity of the encoding system, leading to systematic distortions in internal models of the world.

To explore the geometry of efficient codes, a series of experiments were conducted using β-Variational Autoencoders (β-VAEs). By varying model capacity, introducing biases in the training data, and imposing specific task objectives, the research identifies and characterizes key types of geometric distortions, including prototypization (collapsing similar stimuli into prototypical representations), specialization (over-representing frequent stimuli), and orthogonalization (forming task-specific representations). These distortions, which emerge as principled adaptations to RDT constraints, highlight the trade-offs between accuracy and resource allocation in learned representations.

The second line of inquiry investigates the emergence of number sense in β-VAEs trained solely to reconstruct visual scenes. This work demonstrates that a capacity for numerical perception can arise spontaneously under rate-distortion principles, without explicit supervision for counting or numerical tasks. Behavioral and neural analyses reveal key signatures of human number sense, including Weber's law scaling and distinct mechanisms for processing small versus large numerosities. Furthermore, severely limiting model capacity induces deficits in numerical repre-

sentations that parallel those observed in developmental dyscalculia, suggesting a direct link between resource constraints and cognitive impairments.

This thesis provides compelling computational evidence for the unifying role of RDT in understanding the emergence of intelligent behavior. It demonstrates how efficient coding principles manifest in complex, non-linear systems and offers a framework to understand how statistical properties of the environment shape neural representations. By integrating information theory, cognitive science, and artificial intelligence, this work advances our understanding of how resource limitations sculpt cognitive representations and highlights the potential of RDT to explain the emergence of complex cognitive abilities in both biological and artificial systems.