An information theoretic perspective on cognitive systems: Memory and autonomy

Many biological systems gather and process information about their environment in order to adapt to changing conditions. In the present thesis two fundamental properties, memory and autonomy, underlying this ability are investigated. Both notions are formalized using a general information theoretic framework where the dynamics of a system and its environment are described in terms of observables (random variables) characterizing the state of the system and environment over time.

Memory then refers to the observation that the present behavior can depend on past experiences and is formalized as the mutual information between the system state and previous environmental inputs. Already at this general level, a close relation to dissipative dynamics is established. Further connections to dynamical properties are found in recurrent neural network. Using a mean-field theory the dynamics of randomly connected, input-driven recurrent neural networks can be classified as ordered, i.e. initial state differences are washed out by the input signal (fading memory), or chaotic, i.e. small initial state differences are amplified by the system dynamics. In computer simulations the longest memory as well as most sophisticated information processing capabilities are found in networks close to the transition line from ordered to chaotic dynamics: “computation at the edge of chaos”.

Within the same information theoretic framework *autonomy* is then formalized as dependencies between successive system states which cannot be accounted for by external influences. This captures the idea that the behavior of an autonomous system should not be determined from the outside, but rely upon internal choices and decisions. Interpreting non-trivial memory, i.e. the system has information about past inputs, but no information flow from the environment into the system can be observed, as a general notion of modeling, the following relations between memory and autonomy are established:

- Environmental regularities that are modeled by the system have to be reflected as internal regularities of the system.
- Autonomy and non-trivial memory complement each other since system regularities can either increase autonomy or provide information about the environment.

To further illustrate the proposed measures for autonomy simple automata are used as example systems. Overall, the presented concepts also applies to biological systems, where it can potentially lead to a better understanding of their autonomy and cognition, as well as artificial systems, where it can for example guide the design of autonomous robots.