Subjective randomness and cognitive complexity

Thomas L. Griffiths and Joshua B. Tenenbaum
{gruffydd, jbt}@mit.edu
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Many of the most interesting aspects of human cognition can be described as problems of induction. Learning a language, inferring another person's intentions, and identifying a causal relationship all involve making inferences about the structure of the world from very limited data. In order to be able to solve these inductive problems, we need to have some kind of constraints on the solutions we are willing to consider. Referring to both mathematical results based upon Kolmogorov complexity and a range of psychological studies, Chater and Vitanyi (2003) argue that simplicity may be one of the most important of these constraints.

While simplicity may play an important role in human inference, exploring this idea will require developing cognitive measures of complexity that are more restricted than Kolmogorov complexity. In addition to the practical problem of incomputability, there is a theoretical problem with Kolmogorov complexity: it cannot naturally express the rich prior knowledge that guides human induction. Defining regularities in terms of all programs for a universal Turing machine provides a firm foundation for the study of simplicity, and allows us to answer "asymptotic" questions about learnability (such as why anything is learnable at all), but many of the questions that arise in cognitive science are actually at the opposite extreme: how do we learn so much from so little? People are able to make rapid inferences because their knowledge about the world leads them to have strong biases, considering only a subset of the regularities identified by a Turing machine. The binary sequences HHTHT and HHHHH differ very little in Kolmogorov complexity, but a great deal in human assessments of randomness, while a random sequence of length 1000 and a sequence specified by the parity of the first 1000 digits of π differ a great deal in Kolmogorov complexity, but very little in the responses that people make to them. People have a clear idea of what randomness looks like, and are able to rapidly infer whether a sequence seems to come from a random process.

Understanding the role that simplicity plays in human cognition will require developing accounts of complexity that are easily computable and can incorporate strong inductive biases. We will explore how this might be done, using a statistical approach informed by ideas of complexity to explain people's intuitions about chance. The mathematical definition of randomness has been one of the great successes of Kolmogorov complexity, inspiring several psychological theories (Chater, 1999; Falk & Konold, 1997), making subjective randomness an excellent domain for the development of simpler measures of complexity. We will analyze subjective randomness in terms of an inference as to whether an observation was produced by chance or by some more regular process, implicitly defining a measure of complexity by specifying this regular process. Characterizing the regular process using algorithmic probability or the universal semimeasure leads to an assessment of randomness consistent with the standard treatment in terms of Kolmogorov complexity; choosing other distributions allows us to explore other measures of complexity.

We will use two strategies to define more restrictive cognitive measures of complexity, inspired by algorithmic probability and the universal semimeasure respectively. The algorithmic probability of an object is defined in terms of the shortest program for a universal Turing machine that produces that object and halts. This suggests a method of defining complexity in terms of the length of "programs" for simpler computing machines. Our first strategy for developing more restricted accounts of complexity involves constructing probability distributions corresponding to machines at the first three levels of Chomsky's (1956) hierarchy, using probability to impose a measure of length. We will show that this approach allows us to develop a set of models which vary in their expressive power but provide a good account of perceived randomness in binary sequences. The universal semimeasure is defined as a mixture over all computable probability distributions. Our second strategy defines a restricted form of complexity by taking a mixture of a strict subset of these distributions, characterizing the regular generating process in terms of this mixture. We will show that this approach gives a good account of people's generation of random sequences and numbers, and their assessments of how coincidental they find sets of birthdays and arrays of dots representing bomb locations.

Applying computational measures of complexity in cognitive science will require going beyond universal measures that can be applied in any context. Human beings are able to solve problems of induction because we use rich prior knowledge to constrain the set of possibilities under consideration. However, simplicity still plays a role within this set of possibilities. By considering more restricted measures of cognitive complexity, we are able to show that many human inferences can be characterized in terms of an evaluation of the simplicity of a stimulus.