Compression and intelligence: social environments and communication

David L. Dowe[1], José Hernández Orallo[2], Paramjit K. Das[1]

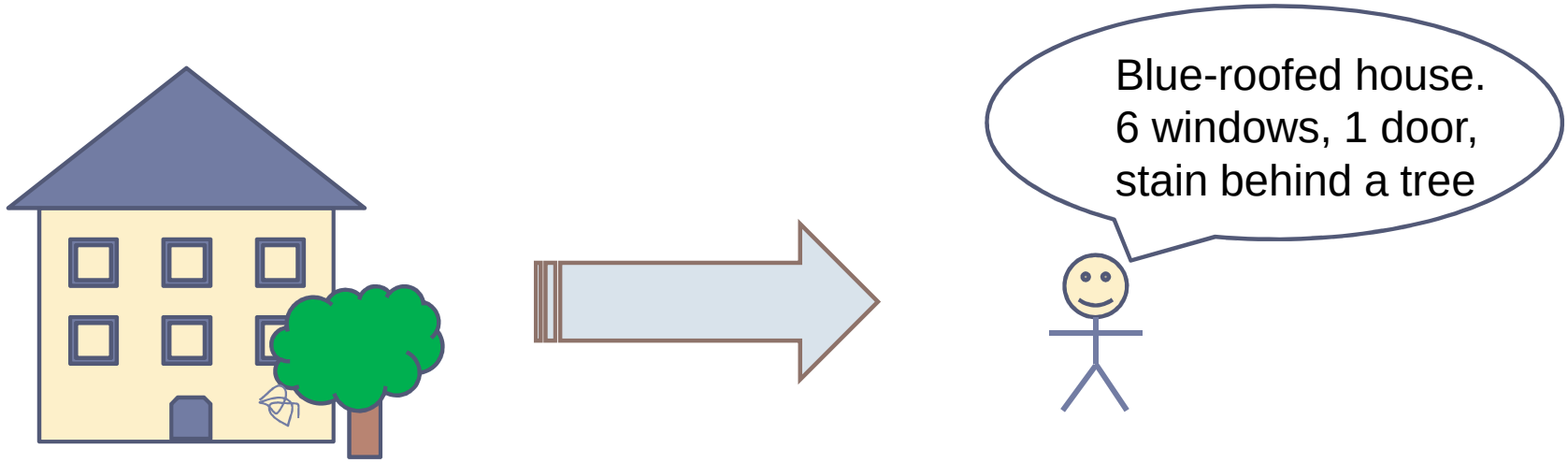
1. *Computer Science & Software Engineering, Clayton School of I.T., Monash University, Clayton, Victoria, 3800, Australia.*
2. *Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València, Spain.*

AGI'2011 - Mountain View, California, 3-6 August 2011.

Outline

- Compression, inference, prediction and intelligence
- Social environments and communication
- Detecting and assessing intelligence
- Conclusions

Compression, inference, prediction and intelligence



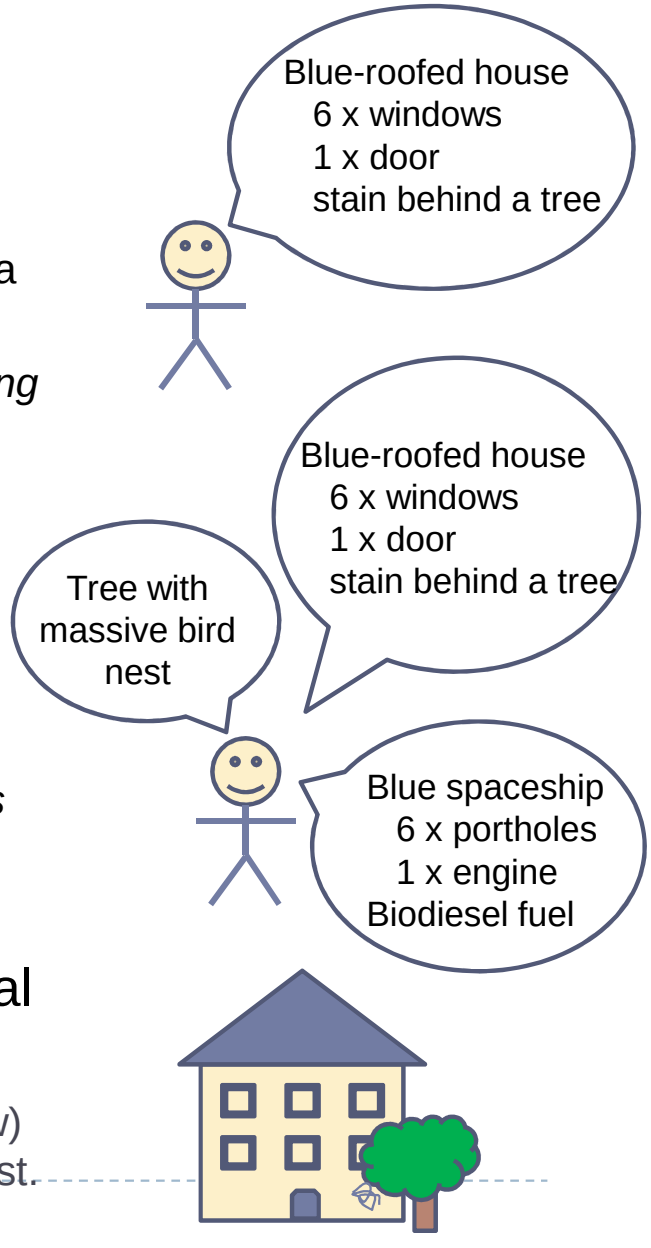
- The relevance of compression to intelligence has been suggested by many.
  - In the last two decades we have seen many intelligence definitions, tests, prizes, etc., based on compression or related ideas.
- But we know that intelligence is not *exactly* compression.
  - Many compression algorithms are able to compress data in a much better way than humans (either lossless or lossy compression).
  - Humans are still better at compressing information which is relevant to their goals or interests.

Compression, inference, prediction and intelligence

- Compression can be seen in many different ways in the context of inductive inference, prediction and intelligence:
  - *One model* (MML inference/explanation) vs. *Many models* (Solomonoff's prediction).
  - *One-part* compression vs. *Two-part* compression.
  - *Lossless* compression vs. *Lossy* compression.

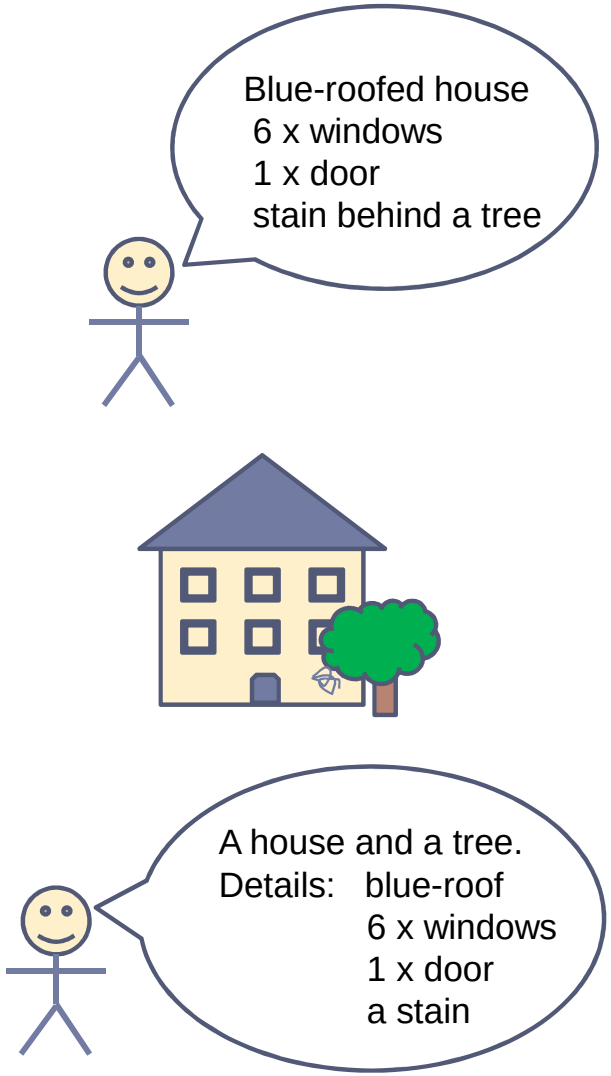
Compression, inference, prediction and intelligence

- One model vs. Many models
  - One model.
    - Minimum Message Length (MML) (Wallace & Boulton 1968) is a common formulation of the idea, with many applications.
    - Caveat: *The best model according to MML might have competing models of similar complexity.*
  - Many models (posterior-weighted mixture of all).
    - Solomonoff's prediction theory (Solomonoff 1964) is the most well-known formulation of the idea, with important results and applications.
    - Caveat: *A (Bayesian) mixture of models (even if weighted by its universal distribution) does not compress the data at all.*
- Solomonoff's approach clearly predicts better in general (even if only slightly) over one single model.
  - But there are many practical advantages of using one (or just a few) models, most especially if there is a model which dominates the rest.



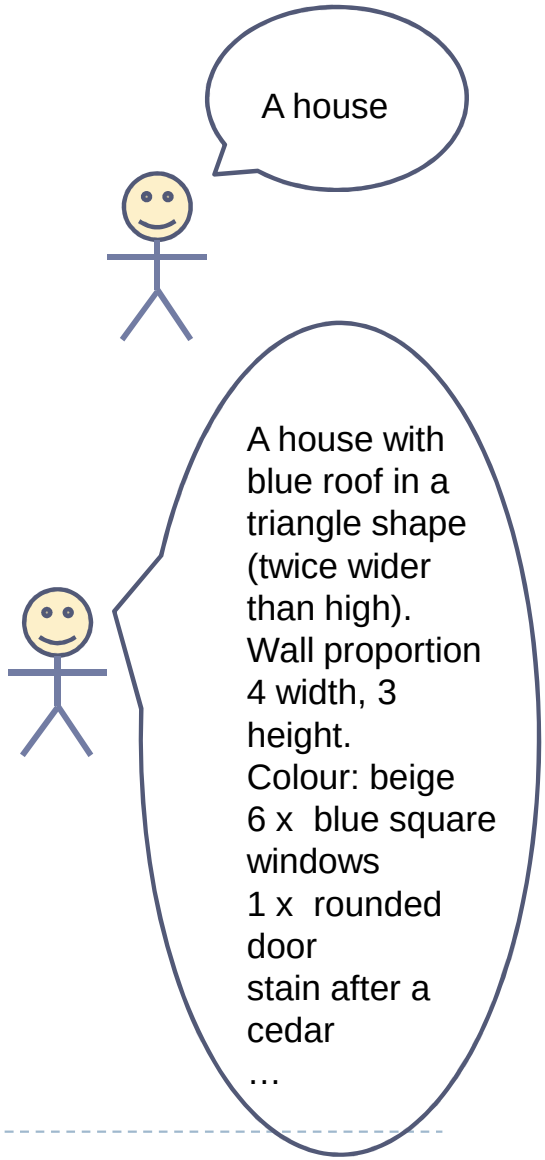
Compression, inference, prediction and intelligence

- One-part vs. Two-part compression
  - In one-part compression, we simply wish to encode the data.
    - We do not care about how intricate the description or code is, if it just compresses the data.
    - Caveat: *One-part compression makes analysis and re-use of 'models' difficult. We don't even talk about "models".*
  - In two-part compression (as MML does), we distinguish between the main pattern and the application of the pattern to encode the data or to add the exceptions.
    - This allows for the identification of the pattern and its reuse for other data and situations.
    - Comment: *The distinction between the two parts is not always unique (in this case we take the one with shortest length).*



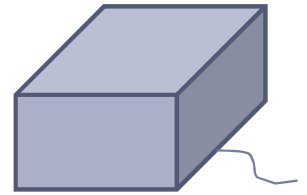
Compression, inference, prediction and intelligence

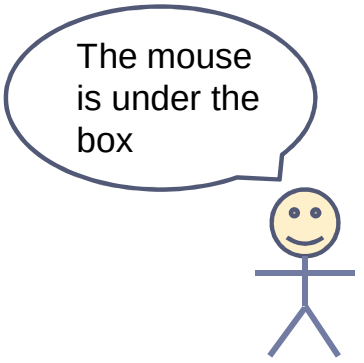
- Lossless vs. Lossy compression
  - Lossy compression is much more common in the real-world.
    - It is more difficult to evaluate since it depends on what part of the data is relevant and what precision is required (*distortion criterion*).
      - Some reinforcement learning systems try to maximise compression in relation to the reward function (the reward is predicted and not the observations).
  - The use of two-part codes implies that the distinction between lossless and lossy compression is more subtle.
    - The main pattern (first part of the message) can represent a lossy (approximated) concept and the second part of the message can equally represent the precision or exceptions.



Social environments and communication

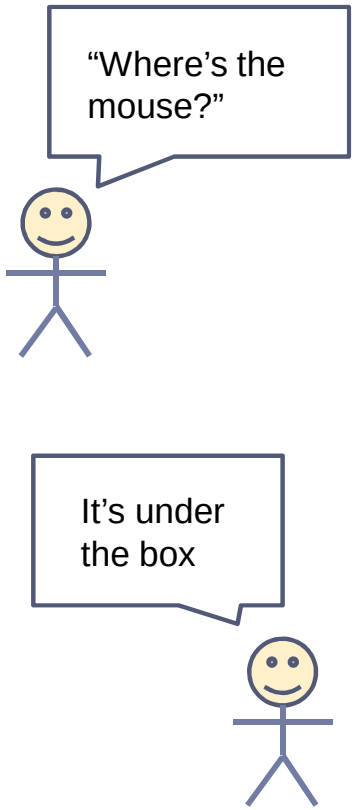
- Competition:
  - The use of a large mixture of models to explain the behaviour of other agents might be optimal in terms of prediction, but it seems inefficient and unrealistic.
  - Mind-reading (between predator and prey, seller and buyer, game opponents, etc.) typically considers a small subset of possible situations and mind states.





Social environments and communication

- Co-operation:
  - Need of shared ontologies, intentions and facts.
  - Use of a single dominant model, and not with many.
- Language:
  - The agents isolate model from data (two-part compression), and are able to communicate the first part (the model) with just the necessary detail.
  - Language is all about sharing concise models, and words are basic units for ("lossily") compressing the world.



Detecting and assessing intelligence

- Introspectively: compression tests have been advocated as a way of detecting and assessing intelligence.
  - Compression-extended Turing tests (Dowe & Hajek 1997a-b, 1998).
  - Measuring the size of the code (compression tests, e.g. Hutter's prize).
  - In general, this is difficult, since the inner knowledge representation may not be accessible, even with the use of language.

- Behaviourally: evaluate the behaviour (or predictability of the models) rather than the models themselves.
  - Some of these approaches use Kolmogorov complexity, universal distributions, etc. (Hernandez-Orallo 1998, Legg & Hutter 2007)
  - The notion of compression is still *implicitly* here:
    - Prediction and compression are related.
    - The complexity of tasks and environments can be assessed by a variant of K().
    - The distribution of tasks may be based on a universal distribution.

Conclusions

- Compression has a fundamental role in intelligence,
  - But the idea of “intelligence as compression” is perhaps too simplistic.
- The issues of one-part vs. two-part, one model vs. many models and lossless vs. lossy compression are very important
  - They must be taken into account and properly specified when talking about compression.
- In *social* environments:
  - One single model can be shared more easily (than multiple models).
  - Two-part (MML) is preferable over one-part to isolate the concept.
  - Lossy compression is much more useful for (concise) communication.