Formal representation, computer implementation, and empirical validation: reflections about the methodology of what we are doing

Benedikt Löwe

Half-day workshop on artificial intelligence, logic, and epistemic planning
DTU, Kungens Lyngby
3 October 2013, 11:45–12:30

Epistemic planning

B. Löwe, E. Pacuit, A. Witzel, DEL planning and some tractable cases, in: H. van Ditmarsch, J. Lang, S. Ju (eds.), Logic, Rationality, and Interaction, Third International Workshop, LORI 2011, Guangzhou, China, October 10–13, 2011. Proceedings, Heidelberg 2011, pp. 179-192 [Lecture Notes in Artificial Intelligence 6953]

T. Bolander and M. B. Andersen, Epistemic planning for single- and multi-agent systems, Journal of Applied Non-Classical Logics 21:1 (2011), pp. 9–34

Modelling (1).

- Abstract away information that we consider irrelevant.
- Represent in terms of a mathematical or formal model. **Translation**
- Implement.
- Phenomenology.
- Make predictions.

Real World $\rightarrow$ Mathematics

Modelling (2).

The *model* and the *modelled* have very different forms of access (in the example: computation vs. measurement).

The *modelled* produces objective data that are easy to interpret, but for some reason are difficult to access.

The *model* produces data of which we are not sure whether it corresponds to reality, but is relatively easy to access.

P. Duhem, La Théorie physique. Son objet et sa structure, 1906.

Epistemic logic (1).

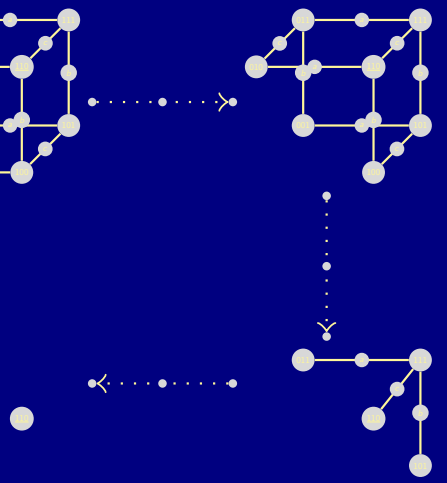
Three brilliant children go to the park to play. When their father comes to find them, he sees that two of them have mud on their foreheads. He then says, "At least one of you has mud on your forehead", and then asks, "Do you know if you have mud on your forehead?" The children simultaneously respond, "No".

The father repeats his question, "Do you know if you have mud on your forehead?" and this time the two children with muddy foreheads simultaneously answer, "Yes, I have mud on my forehead!" while the remaining child answers, "No, I don't know".

Epistemic logic (2).

Semantics of dynamic epistemic logic.

$$
\begin{array}{rll}
\mathbf{M}^V, v \models \mathrm{p}_n & \text{iff} & v \in V(n) \\
\mathbf{M}^V, v \models \mathbf{K}_x\varphi & \text{iff} & \forall w(vRw \to \mathbf{M}^V, w \models \varphi) \\
\mathbf{M}^V, v \models [(\mathbf{S}, s)]\varphi & \text{iff} & \mathbf{M}^V, v \models \mathrm{pre}(s) \\
 & & \quad \text{implies } \mathbf{M} \otimes \mathbf{S}, (v, s) \models \varphi \\
\mathbf{M}^V \models \varphi & \text{iff} & \forall v(\mathbf{M}^V, v \models \varphi) \\
\mathbf{M} \models \varphi & \text{iff} & \forall V(\mathbf{M}^V \models \varphi)
\end{array}
$$

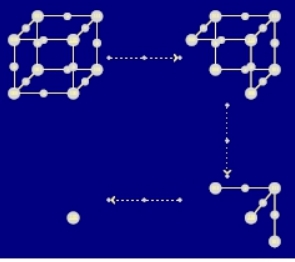


Epistemic logic (3).

Three brilliant children go to the park to play. When their father comes to find them, he sees that two of them have mud on their foreheads. He then says, "At least one of you has mud on your forehead", and then asks, "Do you know if you have mud on your forehead?" The children simultaneously respond, "No".

The father repeats his question, "Do you know if you have mud on your forehead?" and this time the two children with muddy foreheads simultaneously answer, "Yes, I have mud on my forehead!" while the remaining child answers, "No, I don't know".

$$
\begin{array}{rcl}
\mathbf{M}^V, v \models \mathrm{p}_n & \text{iff} & v \in V(n) \\
\mathbf{M}^V, v \models \mathrm{K}_x\varphi & \text{iff} & \forall w(vRw \to \mathbf{M}^V, w \models \varphi) \\
\mathbf{M}^V, v \models [(\mathbf{S}, s)]\varphi & \text{iff} & \mathbf{M}^V, v \models \mathrm{pre}(s) \\
 & & \text{implies } \mathbf{M} \otimes \mathbf{S}, (v, s) \models \varphi \\
\mathbf{M}^V \models \varphi & \text{iff} & \forall v(\mathbf{M}^V, v \models \varphi) \\
\mathbf{M} \models \varphi & \text{iff} & \forall V(\mathbf{M}^V \models \varphi)
\end{array}
$$



We have informal "real world" data on the left side and replaced it with a formal representation on the right side to explain the story.

Modelling in the social sciences / humanities.

Johann Conrad Dannhauer
1603–1666

Christian Weise
1642–1708

Modelling is a hermeneutic process, starting from hermeneutical presumptions (*in bonam partem interpretari*) and aiming to reach *reflective equilibrium*.

Modelling in philosophy: the case of knowledge

Is knowledge justified true belief?

E. Gettier, Is Justified True Belief Knowledge?, Analysis 23 (1963), pp. 121–123

J. Weinberg, S. Nichols, S. Stich, Normativity and Epistemic Intuitions, Philosophical Topics 29 (2001), pp. 429-460.

Back to epistemic planning

- A finite number of facts p_0, ..., p_n.
- Initial knowledge states for all agents.
- A list of possible actions:
  - Ask whether p_i.
  - Refuse to answer.
  - Answer that you don't know.
  - Answer truthfully.
  - Publicly announce the answer.

Epistemic planning amounts to finding an algorithm that produces a sequence of actions to produce a given epistemic state of the agents, e.g.,

$$(\Box_0 p \vee \Box_0 \neg p) \wedge \neg \Box_1 p \wedge \neg \Box_1 \neg p.$$

Some references relating to this morning's discussion

D. O. Stahl, P. W. Wilson, On players' models of other players: Theory and experimental evidence, *Games and Economic Behavior* 10 (1995), pp. 218–254

B. Keysar, S. Lin, D. J. Barr, Limits on theory of mind use in adults, *Cognition* 89 (2003), pp. 25–41

B. Meijering, H. van Rijn, N. A. Taatgen, and R. Verbrugge, I do know what you think I think: Second-order theory of mind in strategic games is not that difficult. in: L. Carlson, C. Hölscher, and T. Shipley (eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Cognitive Science Society, Austin, TX, 2011, pp. 2486-2491.

B. Löwe, E. Pacuit, S. Saraf, Identifying the structure of a narrative via an agent-based logic of preferences and beliefs: Formalizations of episodes from CSI: Crime Scene Investigation^{TM}, in: M. Duvigneau, D. Moldt (eds.), Proceedings of the Fifth International Workshop on Modelling of Objects, Components and Agents, MOCA'09, Hamburg 2009 [FBI-HH-B-290/09], pp. 45-63