Representing Plausible Beliefs about States, Actions, and Processes

Toryn Qwyllyn Klassen

Department of Computer Science
University of Toronto

November 9, 2020

Introduction

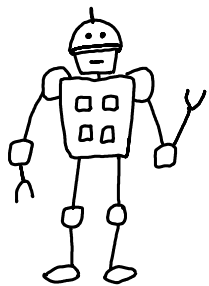
The topic is modelling beliefs about a dynamic world in a way that allows for changes based on observations made by the agent.

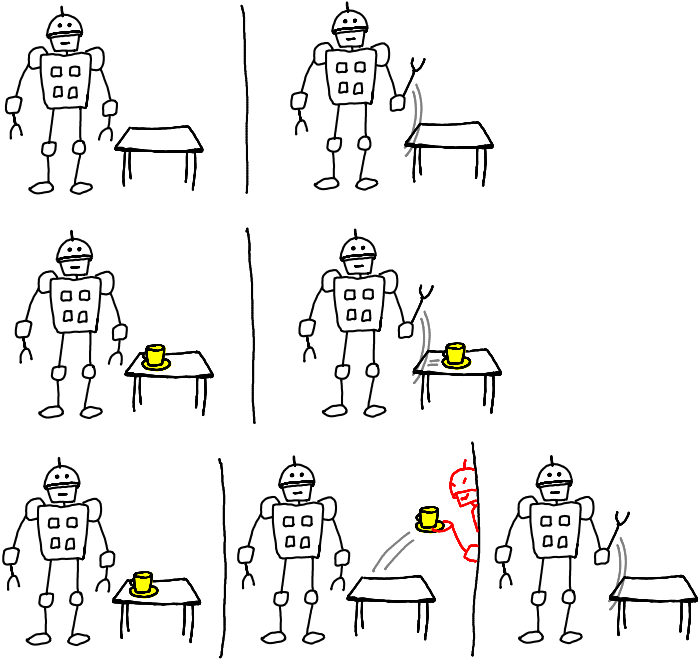
The thesis is concerned with representing, within a logical theory,

1. what initial conditions the agent considers (im)plausible,
2. what effects the agent thinks actions (im)plausibly have,
3. and what sequences of actions the agent thinks have (im)plausibly occurred or will occur.

Representing plausibility supports modelling changing beliefs, since when the most plausible options are refuted by observations, the agent can fall back to the next most plausible options.

- Imagine an agent that believes the coffee cup is on the table.
- The agent then moves its arm and hand and believes that it has picked up the cup.
- However, the agent then senses that its hand is empty.





The situation calculus

Reiter's version of the situation calculus is a language in second-order logic, in which situations represent histories of actions.

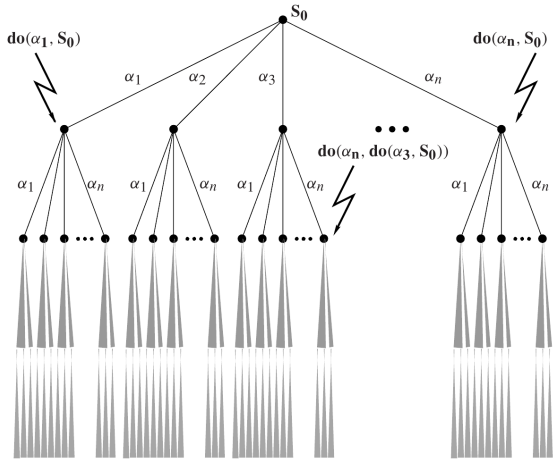


Image copied from Reiter (2001, Figure 4.1).

The situation calculus

- Properties that can vary between situations are represented using **fluents**, predicates that take a situation argument.
  - For example, Holding(x, y, s) might represent whether x is holding object y in situation s.
- An environment can be described in the situation calculus with a set of axioms, an **action theory**.
- Action theories traditionally include axioms describing
  - the initial state,
  - the preconditions of actions,
  - and how each fluent is changed by actions.

Modelling belief

- The standard way of describing beliefs or knowledge in logic is in terms of **possible worlds**.
- An **accessibility relation** relates world w to world v if in w the agent considers that v may be the actual world.
- What is known or believed by an agent in a particular world is defined as what is true in **all accessible possible worlds**.
- Belief and knowledge can be described in **modal** logics that introduce special operators for these modalities.
- Alternatively, an accessibility relation can be encoded in classical **first-order** (or **second-order**) logic.
  - Scherl and Levesque (2003) did this in the situation calculus, using situations as the "possible worlds".

Plausibility

- To specify how beliefs can change and be retracted over time, further structure beyond the possible worlds model is needed.
- An ordering on possible worlds can be used to govern how beliefs get changed.
  - The agent's new beliefs can be determined by the best worlds (according to the ordering) that are consistent with the new information (Grove, 1988).
- The ordering can be thought of as indicating which worlds are more **plausible** than others.
- Shapiro et al. (2011) defined belief in the situation calculus as what is true in all the **most plausible** accessible situations.

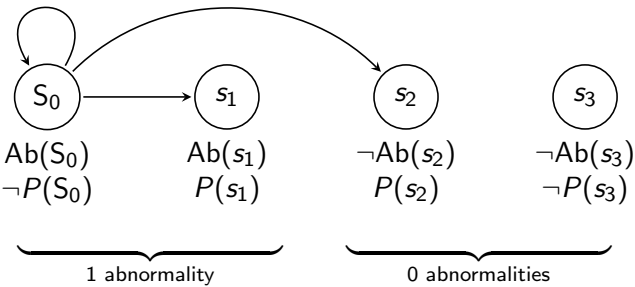
Thesis statement

Measuring the plausibility of a situation by counting the number of abnormalities contained within it allows for a perspicuous way of representing revisable beliefs about various aspects of a dynamic environment, including its state, the effects and preconditions of actions, and the behavior of environment processes.

Chapter 3: Specifying plausibility levels

- This chapter extends Shapiro et al.'s approach by measuring plausibility by counting the number of **abnormal** atomic formulas true in a situation.
- This is related to **cardinality-based circumscription** (Liberatore and Schaerf, 1997; Sharma and Colomb, 1997; Moinard, 2000).

Example



- The accessible situations (from S_0) are those situations s in which $\neg \mathrm{Ab}(s) \supset P(s)$ is true.
- The set of most plausible accessible situations is $\{s_2\}$.
- $P(s)$ is true at each most plausible accessible situations s.

Contributions of Chapter 3

- proposes counting abnormalities as a way of defining plausibility levels, and formalizes this using second-order logic
- unlike the rival method for specifying plausibility levels from Schwering and Lakemeyer, allows
  - for features that **independently** contribute to the plausibility of a situation to be easily described
  - for an **infinite** number of plausibility levels to be described
- proves a result on how cardinality-based circumscription generalizes a form of **lexicographic entailment**
- shows how changing abnormalities can be used to assign plausibilities to the (non-)occurrence of exogenous actions
- proves, for action theories that allow the agent to have incorrect knowledge about the effect of actions, how closely they follow the AGM postulates for belief revision

Chapter 4: Specifying the plausibility of domain dynamics

This chapter applies the approach of specifying plausibility levels from the previous chapter to describing the effects of actions, and to model change of beliefs about action effects.

It focuses on how theories can be written so as to control how general of conclusions an agent should draw from observations.

Contributions of Chapter 4

- proves that (in some cases) when the axioms describing domain dynamics are written to refer to abnormalities, the agent will believe "normalized" axioms that don't refer to abnormalities
- proposes **patterns** to follow when writing axioms about action effects, in order to control how general of conclusions the agent draws about the behavior of actions from unexpected observations
- shows how our theories can be used to model changing beliefs about
  - the results of **sensing**
  - the **preconditions** of actions
- describes how to apply **regression** with our theories, including how to use beliefs about action effects within the regression procedure, and proves its correctness

Chapter 5: Environment processes and knowing how

This chapter extends our formal model of the beliefs of an agent to take into account ongoing **exogenous** processes.

It also gives a formalization of **knowing how** to achieve goals in such a setting, defining knowing how in terms of belief.

Contributions of Chapter 5

- presents an approach to modeling defeasible belief in the situation calculus where the accessible situations over time are constrained to be reachable by following a ConGolog program
- proves that under some conditions, if the ConGolog program that's running refers to abnormalities, the agent will believe that a simpler "normalized" program that doesn't refer to abnormalities is running
- introduces a definition of **knowing-how** in terms of belief, that takes into account both how beliefs may be false and the running of exogenous processes
  - proves that this definition **generalizes** Lespérance et al.'s (2000), among other properties
  - also formalizes a version of knowing-how which describes goals that can be achieved with **sequential plans**
  - the approach supports **revision** of beliefs about knowing-how

Summary

- representing plausibility by counting abnormalities (Chapter 3)
- referring to abnormalities in dynamics axioms (Chapter 4)
- programs describing exogenous processes using abnormalities (Chapter 5)

Future work

Some possibilities:

- modelling plausibility in formalisms other than the situation calculus
- considering **belief update** along the lines proposed by Boutilier (1996), where changes to the world are made by unobserved exogenous events
- looking at **elaboration tolerance** – how easy is to change the plausibility ordering induced by a knowledge base?

References

Craig Boutilier. Abduction to plausible causes: an event-based model of belief update. *Artificial Intelligence*, 83(1): 143–166, 1996. doi: 10.1016/0004-3702(94)00097-2.

Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17(2):157–170, 1988. doi: 10.1007/BF00247909.

Yves Lespérance, Hector J. Levesque, Fangzhen Lin, and Richard B. Scherl. Ability and knowing how in the situation calculus. *Studia Logica*, 66(1):165–186, 2000. doi: 10.1023/A:1026761331498.

Paolo Liberatore and Marco Schaerf. Reducing belief revision to circumscription (and vice versa). *Artificial Intelligence*, 93 (1):261–296, 1997. doi: 10.1016/S0004-3702(97)00016-7.

Yves Moinard. Note about cardinality-based circumscription. *Artificial Intelligence*, 119(1):259–273, 2000. doi: 10.1016/S0004-3702(00)00018-7.

Raymond Reiter. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. MIT Press, 2001.

Richard B. Scherl and Hector J. Levesque. Knowledge, action, and the frame problem. *Artificial Intelligence*, 144(1):1–39, 2003. doi: 10.1016/S0004-3702(02)00365-X.

Christoph Schwering and Gerhard Lakemeyer. A semantic account of iterated belief revision in the situation calculus. In *ECAI 2014 - 21st European Conference on Artificial Intelligence*, pages 801–806, 2014.

Steven Shapiro, Maurice Pagnucco, Yves Lespérance, and Hector J. Levesque. Iterated belief change in the situation calculus. *Artificial Intelligence*, 175(1):165–192, 2011. doi: 10.1016/j.artint.2010.04.003.

Nirad Sharma and Robert Colomb. Towards an integrated characterisation of model-based diagnosis and configuration through circumscription policies. Technical Report 364, Department of Computer Science, University of Queensland, 1997.