# Managing Trust and Uncertainty for Distributed AI Systems

M.Ravi

Université Joseph Fourier,
Orange Labs

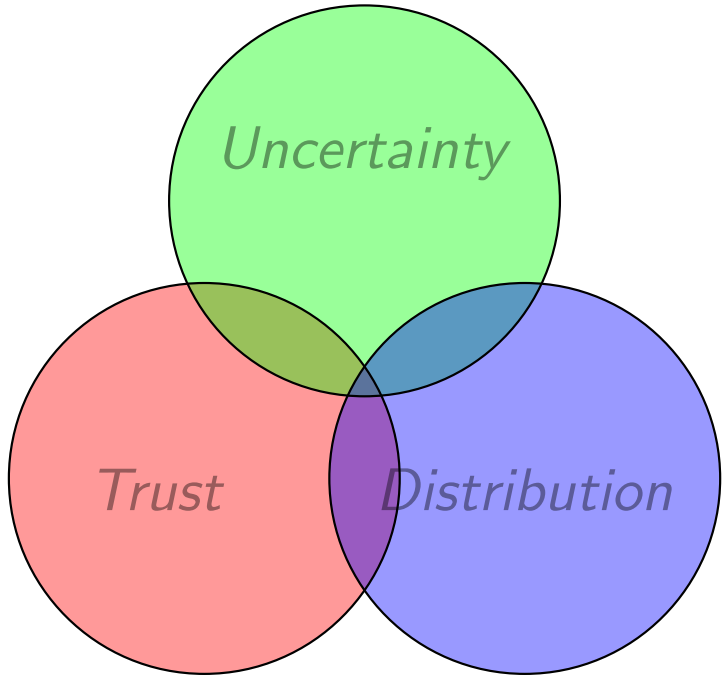
*ravi@imag.fr*

Advisors
Yves Demazeau, MAGMA/LIG, CNRS
Fano Ramparany, Orange Labs, Meylan

June 28, 2014

# Outline

# Introduction



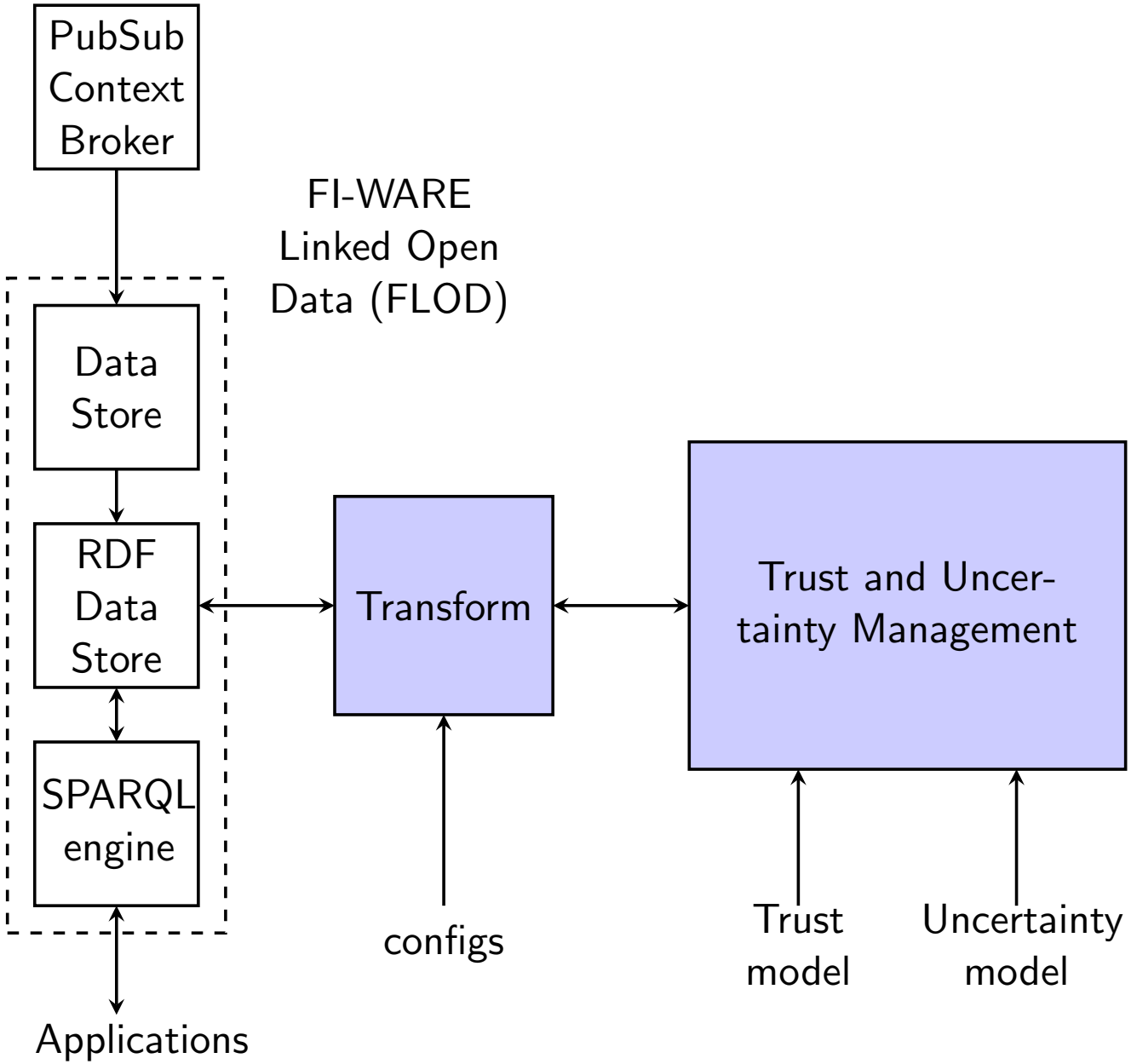

## Problems

1. Survey on existing state of the art in Trust and Uncertainty domains
2. Explore a relationship between Trust and Uncertainty
3. Propose a Multi-agent System architecture for the management of Trust and Uncertainty

# Motivation

1. FI-PPP or Future Internet Public Private Partnership (`http://www.fi-ppp.eu`)
2. Project FI-WARE
   - A common infrastructure which supports the requirements of several sectors like health care, environment, and telecommunications.
   - Includes different partners Telefonica, Orange, Fraunhofer Focus, etc.
3. Management of Trust and Uncertainty - a common requirement of all.

# Project FI-WARE



# Use case

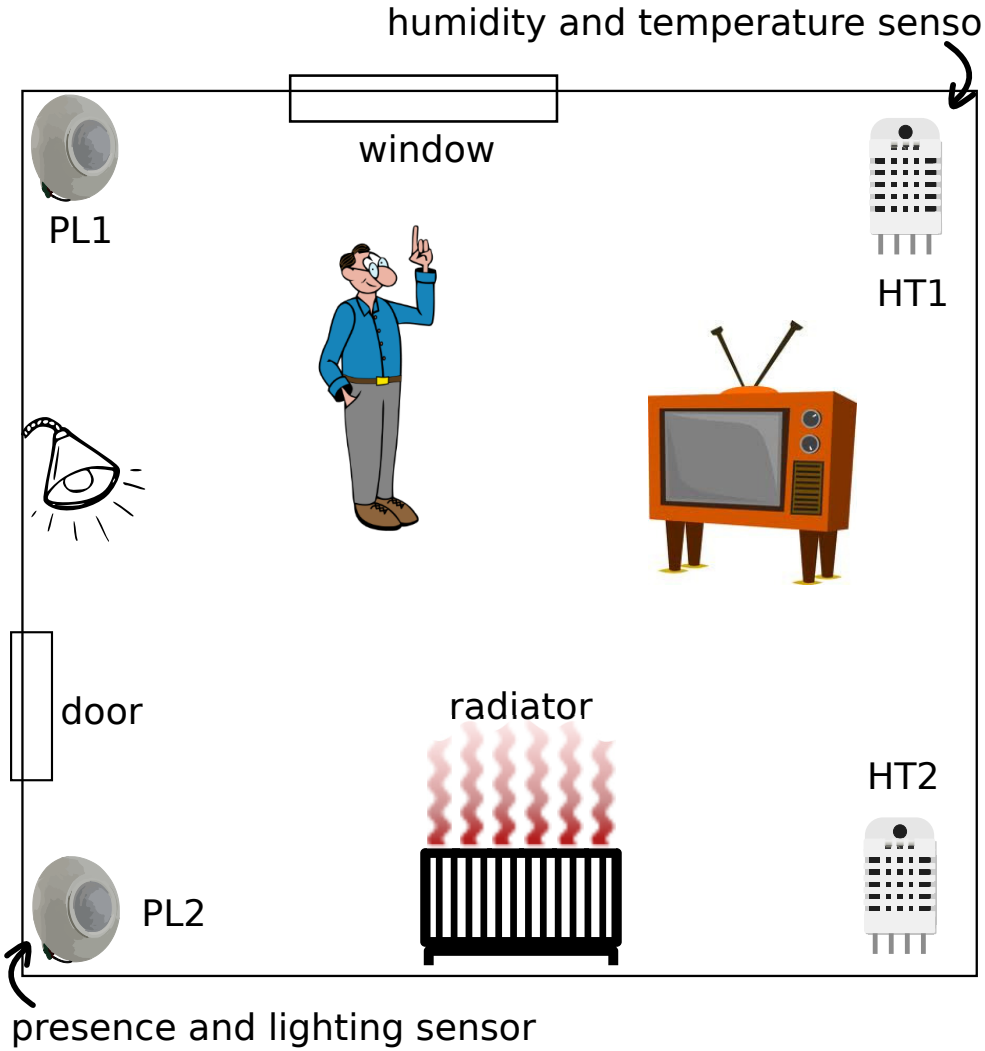
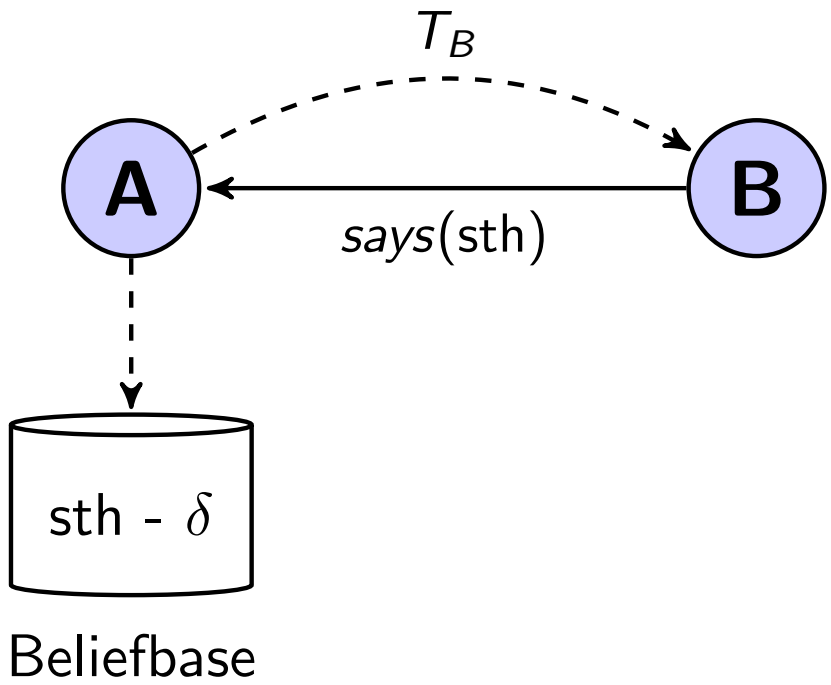



Figure: Smart home with devices (TV, lamp, radiator) controlled by an intelligent system

# Trust and Uncertainty




1. Problem: How to convert $T_B$ to $\delta$?
2. Trust is related to information source and uncertainty is related to the information.

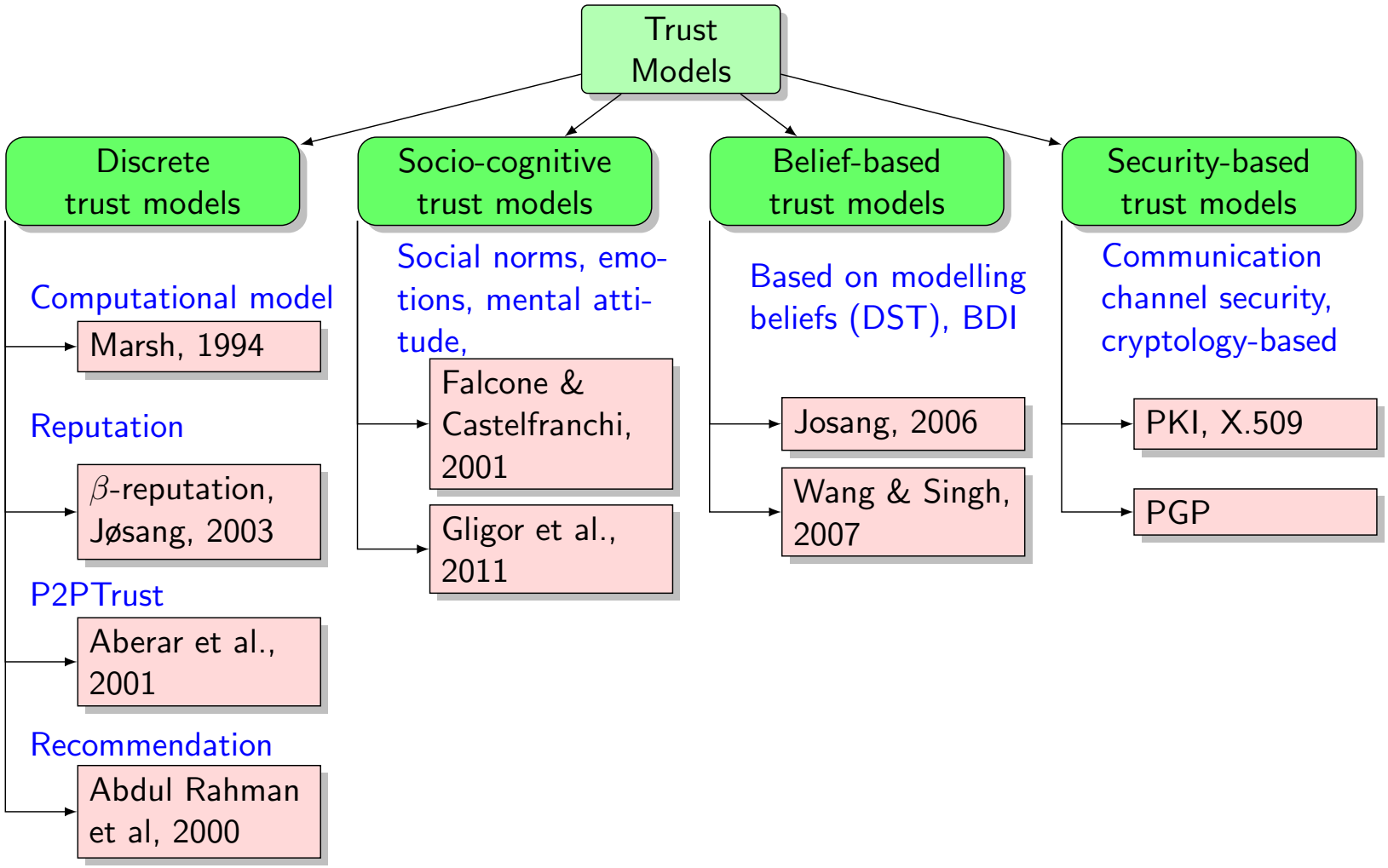
# Trust


- Trust has no specific definition as it varies from person to person and domain to domain.
- It is a cross-domain commodity.

- Golembiewski & McConkie (1975)

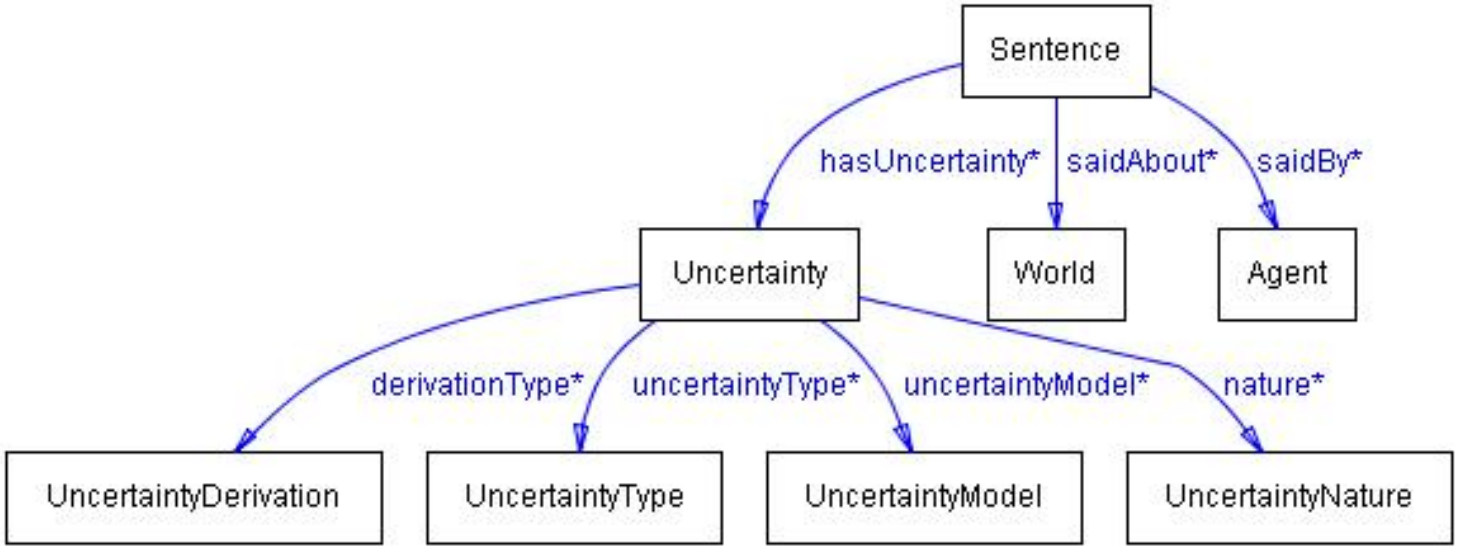
  Trust is strongly related to confidence in something, be it person to be trusted, the environment, or whatever it is that the desirable outcome is contingent upon.

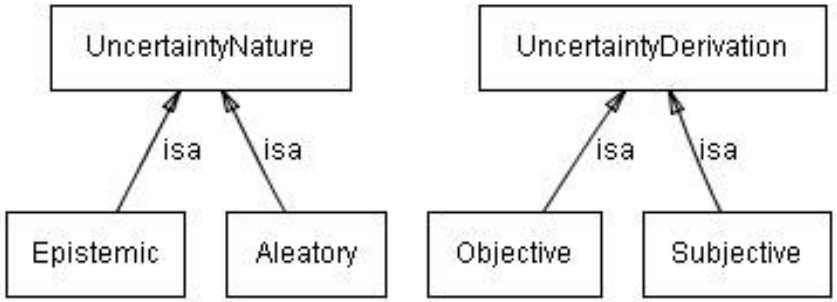
# Trust models



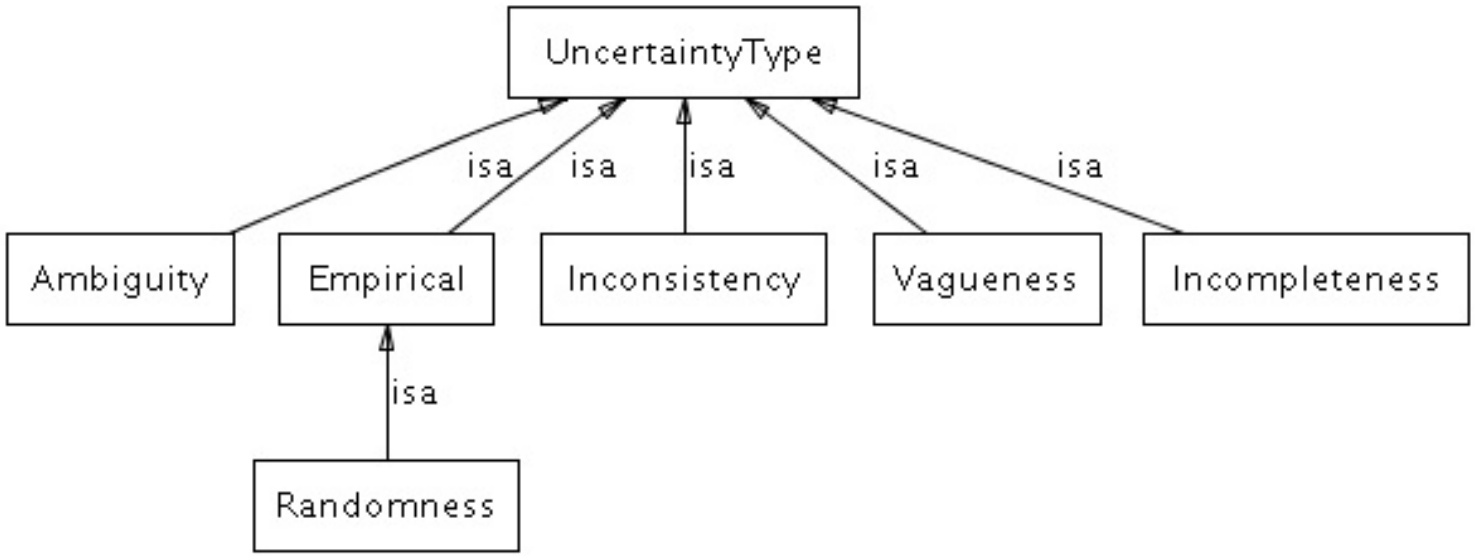
# Uncertainty

Uncertainty Reasoning for the World Wide Web - `http://www.w3.org/2005/Incubator/urw3/XGR-urw3-20080331/`



# Sources and types of uncertainty



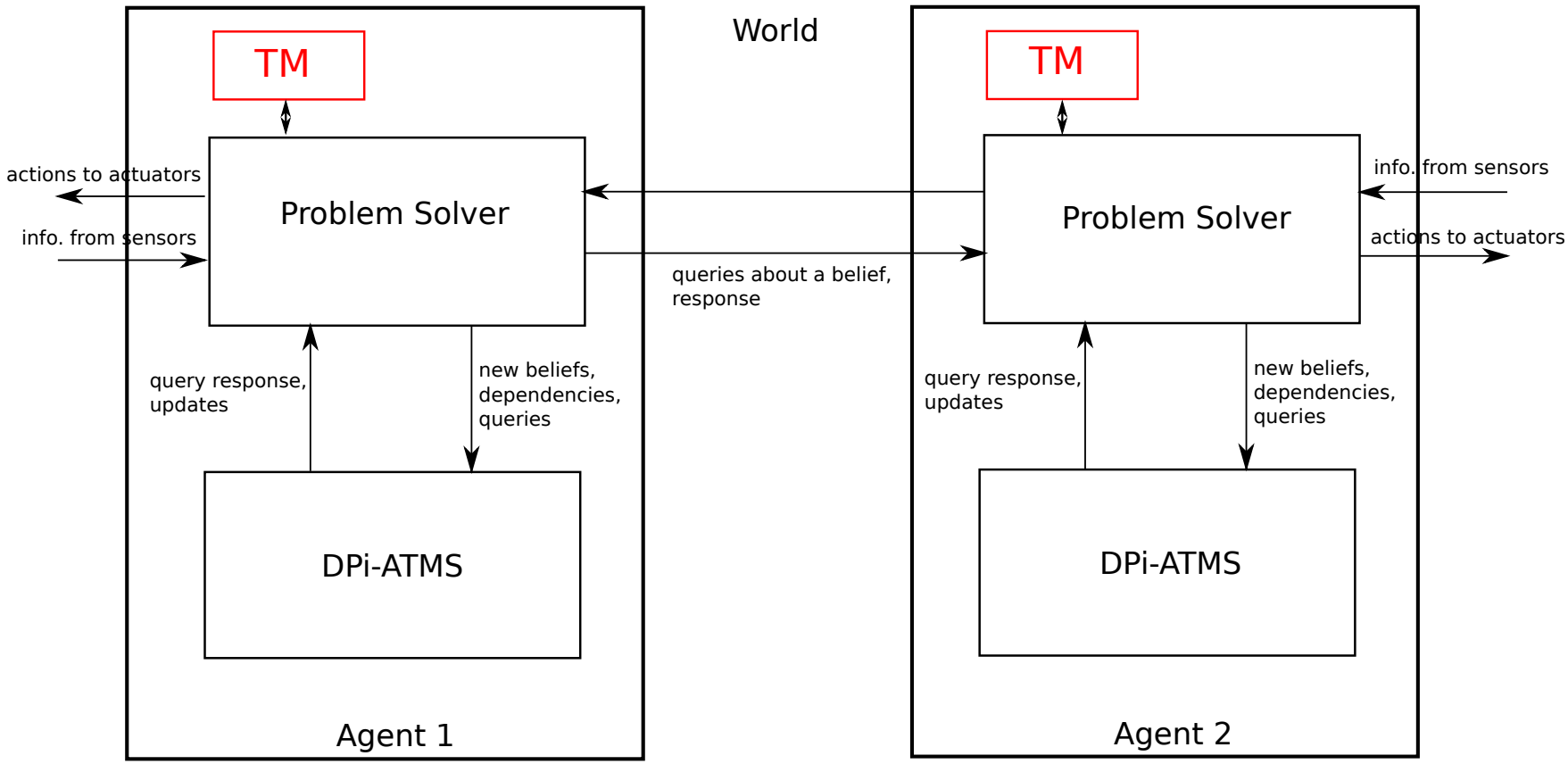

## Sources of uncertainty

- Inaccurate sensors
- Untrustworthy data source
- Abductive reasoning
- Chain uncertainty
- Absence of sufficient data

## Modeling uncertainty

- Possibilistic Logic
- Probabilistic Logic
- Fuzzy Logic
- Dempster-Shafer Theory
- Subjective Logic

# Proposed architecture



# ATMS

1. What is ATMS?
   - Assumption-based Truth Maintenance System (de Kleer, 1986)
   - A belief-revision system
2. Basic terms
   - Premise
   - Assumption
   - Environment: An environment $E$ of a node $n$ is a set of assumptions ($E = a_1, a_2, ...a_i$) the conjunction of which derives the node, i.e., $a_1 \wedge a_2 \wedge ...a_i \rightarrow n$
   - Label: A label $L$ is a set of environments for a given node $n$
   - Justification: A justification $J$ relates how a node can be derived from any other node(s)

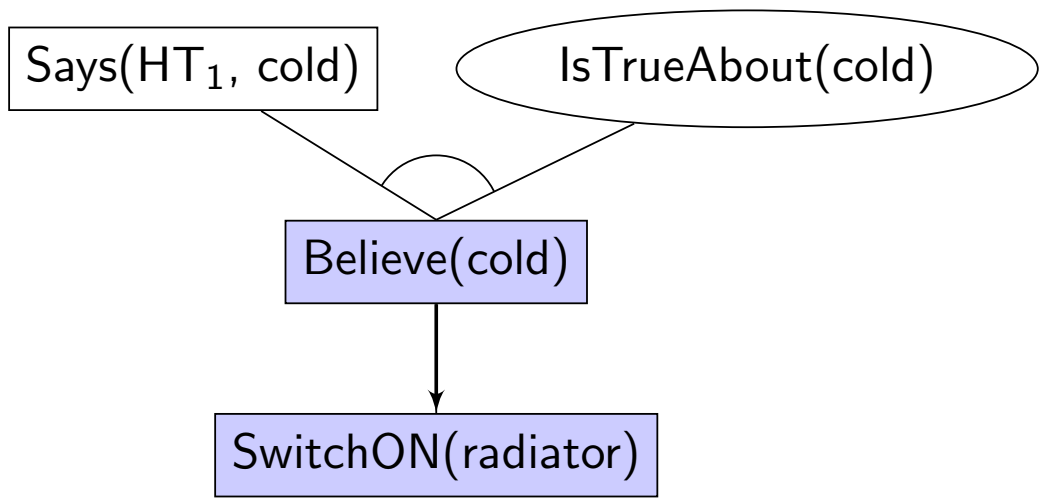
# Working of ATMS

$$Says(X, sth.) \wedge IsTrueAbout(X, sth.) \rightarrow Believe(sth.) \quad (1)$$

$$Believe(cold) \rightarrow SwitchON(radiator) \quad (2)$$

$$IsTrueAbout(HT_1, coldweather) \quad (3)$$

$$Says(HT_1, cold) \quad (4)$$



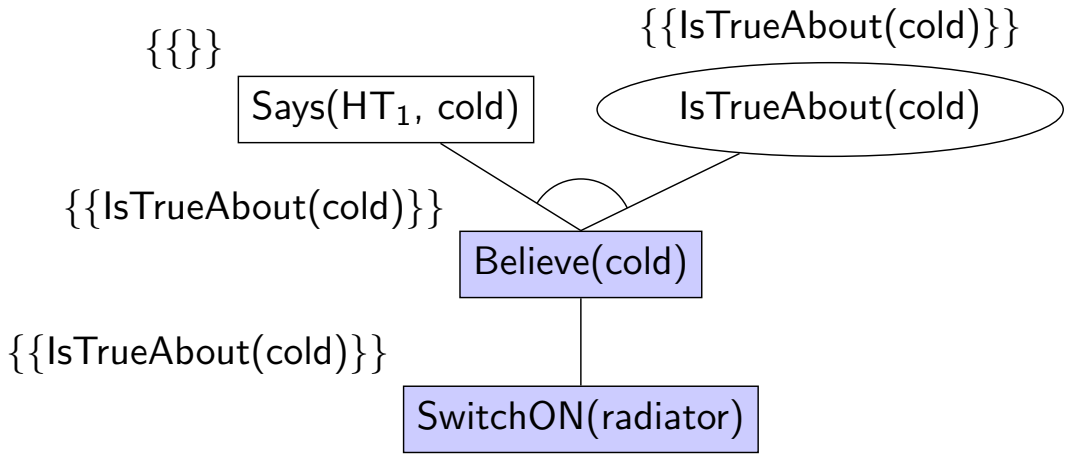
# Working of ATMS

$$Says(X, sth.) \wedge IsTrueAbout(X, sth.) \rightarrow Believe(sth.)$$

$$Believe(cold) \rightarrow SwitchON(radiator)$$

$$IsTrueAbout(HT_1, coldweather)$$

$$Says(HT_1, cold)$$



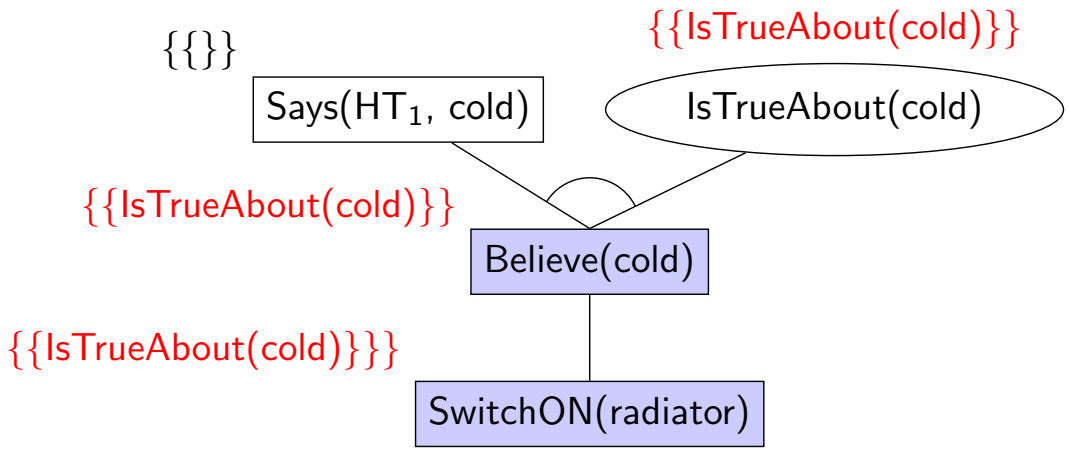
# Belief revision

$$Says(X, sth.) \wedge IsTrueAbout(X, sth.) \rightarrow Believe(sth.)$$

$$Believe(cold) \rightarrow SwitchON(radiator)$$

$$Says(HT_1, cold)$$

$$IsTrueAbout(HT_1, coldweather)$$



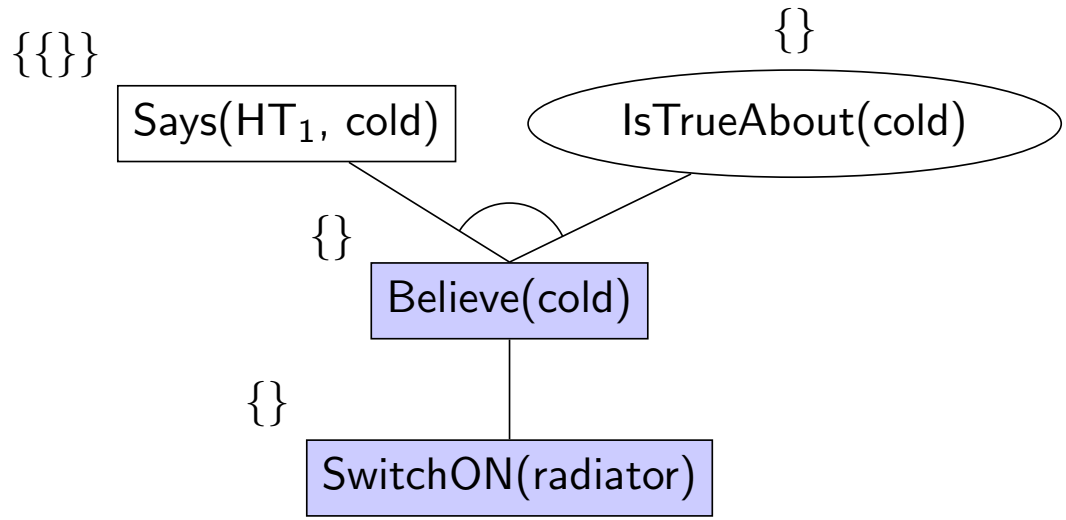
# Belief revision

$$Says(X, sth.) \wedge IsTrueAbout(X, sth.) \rightarrow Believe(sth.)$$

$$Believe(cold) \rightarrow SwitchON(radiator)$$

$$Says(HT_1, cold)$$

$$IsTrueAbout(HT_1, coldweather)$$



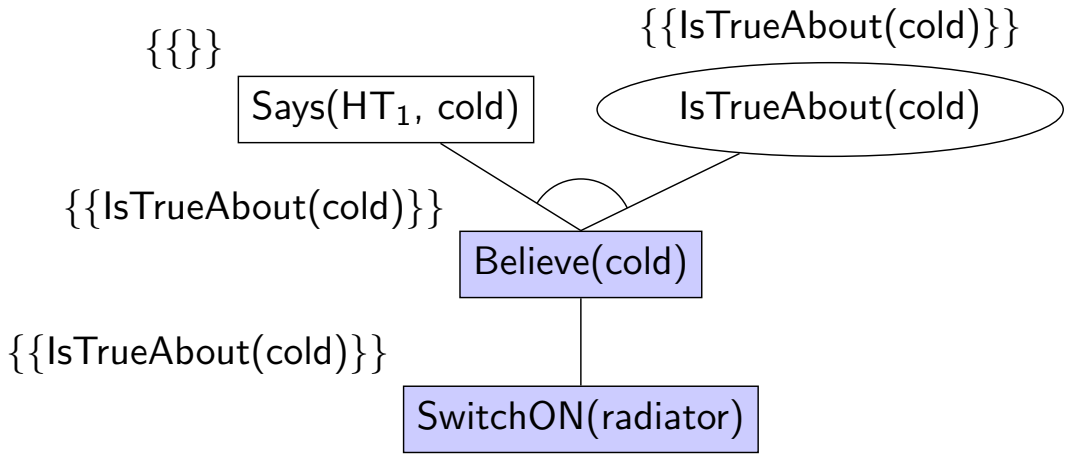
# ATMS to Π-ATMS

$$Says(X, sth.) \wedge IsTrueAbout(X, sth.) \rightarrow Believe(sth.)$$

$$Believe(cold) \rightarrow SwitchON(radiator)$$

$$IsTrueAbout(HT_1, coldweather) \qquad (N{=}0.8)$$

$$Says(HT_1, cold) \qquad (N{=}1)$$



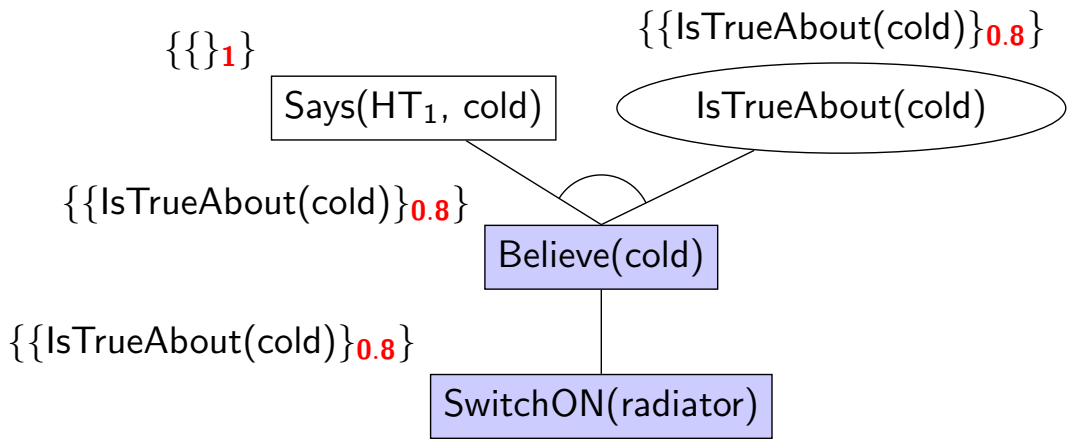
# ATMS to Π-ATMS

$$Says(X, sth.) \wedge IsTrueAbout(X, sth.) \rightarrow Believe(sth.)$$

$$Believe(cold) \rightarrow SwitchON(radiator)$$

$$IsTrueAbout(HT_1, coldweather) \qquad (N{=}0.8)$$

$$Says(HT_1, cold) \qquad (N{=}1)$$



# Conclusions/Discussions

- Trust is subjective. It varies from domain to domain. It is difficult to have a generic model of trust for all domains that includes all aspects.
- ATMS enables an agent to maintain a consistent set of beliefs. When coupled with possibilistic logic (Π-ATMS), the beliefs can be quantified in terms of necessity.
- Conversion of trust to uncertainty is tricky. The hypothesis that there exists proportionality relation between trust and necessity needs validation.

Thank you.

# References


1. A. Abdul-Rahman, and S. Hailes. Supporting trust in virtual communities. *Proceedings of Hawaii International Conference on System Sciences*, 2000.
2. A. Jøsang, and R. Ismail. The beta reputation system. *Proceedings of the 15th bled electronic commerce conference*. 2002.
3. D. Dubois, J. Lang, and H. Prade. A possibilistic assumption-based truth maintenance system with uncertain justifications, and its application to belief revision. *Truth maintenance systems*, 1991.
4. J. de Kleer. An assumption-based TMS. *Artificial Intelligence*, 28(2):127162, Mar. 1986.
5. Y. Wang, and M.P. Singh. Formal Trust Model for Multiagent Systems. *IJCAI*, 2007.