

## Explanations and Knowledge-Based Systems

### ► Kinds and Goals of Explanations

Dr. Thomas Roth-Berghofer  
Knowledge-based Systems Group  
Technische Universität Kaiserslautern  
and  
Knowledge Management Research Department  
German Research Center for Artificial Intelligence DFKI GmbH

## Overview

- 1 Motivation
- 2 Kinds of Explanations
- 3 Explanation Goals

# What is an explanation?

- In Philosophy of Science, the main kind of explanation discussed are **scientific explanations**:
  - Answers to why-questions:
    - Can some fact  $E$  (the *explanandum*) be derived from other facts  $A$  with the help of general laws  $L$  (the *explanans*  $L \cup A$ )?
- Scientific explanations are “domain dependent”:
  - Laws of Nature in Physics vs. “Social Laws” in Social Sciences
- Distinguish between:
  - Cause-giving explanations: Why does an explanandum event occur? What is its cause for being? (‘Seinsgrund’)
  - Reason-giving explanations: Why is it reasonable to believe that the explanandum event has occurred or will occur? (‘Vernunftgrund’)

Q: Why doesn't Grandma visit us anymore?

E: Sweetheart, Grandma is taking a very long journey.



**All of the explanations are literally false!**



Q: Miss Thornton, where do little babies come from?

E: Well, Bernard, when a mom and a dad love each other very much they come together and have a baby.



Q: Why does it appear that in every interaction, the total linear momentum of interacting bodies remains constant?

E:  $F = ma$  (Newton's second law).

For every interaction there is an opposite reaction (Newton's third law).

In every interaction, the total linear momentum of the system of interacting bodies remains constant. (law of conservation of linear momentum)



Q: Could you please tell me why you've been away all night?


E: I'm sorry darling, but there was so much work that my colleagues and me had to stay in the office.



D. Cohnitz. Explanations are like salted peanuts. In Ansgar Beckermann and Christian Nimtz, editors, Fourth International Congress of the Society for Analytic Philosophy, 2000.

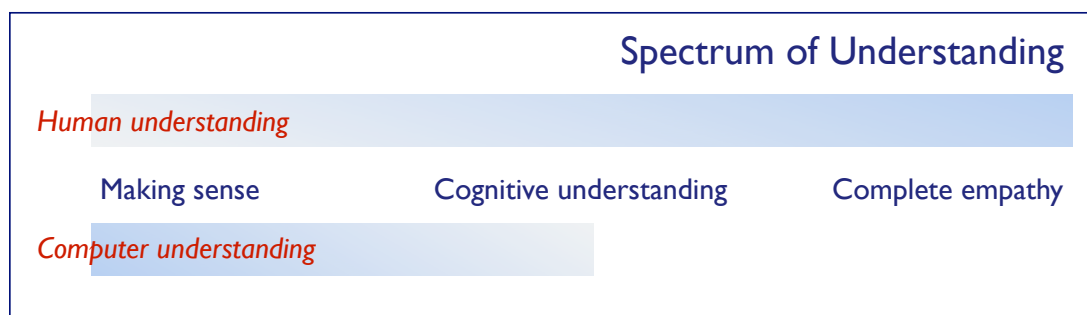
## Explanations mainly support human decision making

- Explain the path that led to a solution.
- Explain, how the respective system (human or machine) works.
- Explain, how to handle the system (human or machine).
- Explain failures.
- **Thus:**  
Explanations need to be **inclusive** as well as **instructive**.

 Roger C. Schank. *Explanation Patterns: Understanding Mechanically and Creatively*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.

## Explaining to oneself means understanding

- As soon as a system explains its own actions not only to those who inquire but to itself, the system becomes an understanding system.
- Computer understanding can only claim the left half of the spectrum, from making sense to cognitive understanding.



## Illustrating Examples

### ■ Making sense

- input: news from a newspaper
- output: a summary of a newspaper story – a translation of a speech into another language

### ■ Cognitive understanding

- input: a set of stories about airplane crashes, complete with data about the airplanes and the circumstances
- output: possible causes of the crashes based on a theory of the physics involved and an understanding of the airplane design, used in conjunction with algorithms that can do creative explanation

### ■ Complete empathy

- output: expressing feelings

## What must be explained?

### ■ Physical world

- how 'things' work

### ■ Social world

- how societies work

### ■ Individual patterns of behavior

- how individuals behave

## Purpose of explaining

- Technical purposes
- Increasing confidence into a system
  - by providing explanations as part of the result
  - by improving the quality of the result
  - by providing evidence of how the output was derived
  - by making the system accountable for its predictions
  - by giving the user a sense of control over the system

## Overview

- 1 Motivation
- 2 Kinds of Explanations
- 3 Explanation Goals

## Useful kinds of explanations

- Conceptual explanations Scientific explanations
  - “What is ...?” or “What is the meaning of ...?”
- Why-explanations
- How-explanations
- Purpose-explanations
  - “What is ... for?” or “What is the purpose of ...?”
- Cognitive explanations



Peter Spieker. Natürlichsprachliche Erklärungen in technischen Expertensystemen. Dissertation, University of Kaiserslautern, 1991.

## Conceptual Explanations

- The goal of conceptual explanations is to build links between unknown and known concepts.
- Examples:
  - Definition:  
“What is a bicycle” – “A bicycle is a land vehicle with two wheels in line. Pedal cycles are powered by a seated human rider and are a form of human powered vehicle.”
  - Theoretical proposition:  
“What is force?” – “Force is Mass times Acceleration.”
  - Prototypical usage of individual things or actions:  
“What is a bicycle?” – “The thing, the man there crashed with.”
  - Functional mapping:  
“What is a bicycle?” – “A bicycle serves as a means of transport.”

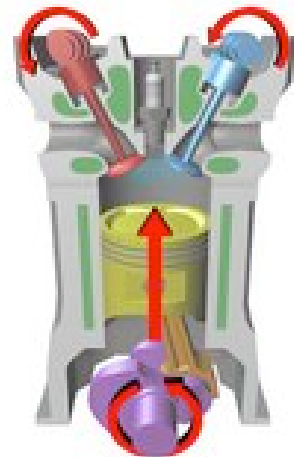


## Why-explanations

- Why-explanations provide causes or justifications for facts or events.
- Causes are never symmetrical, justifications (providing only evidence for the explained) could be symmetrical.
- Examples:
  - Justification:  
“Why does the universe expand?” – “Because we can observe a red shift of the light emitted by other galaxies.”
  - Cause:  
“Because the whole matter was concentrated at one point of the universe and because the whole matter moves away from each other.”

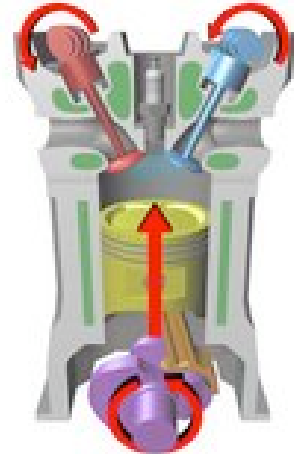
## How-explanations

- The goal of how-explanations is to help the questioner understand the functionality of an object.
- Example:
  - “How does a combustion engine work?”  
– “A combustion engine is an engine that operates by burning its fuel.”



## Purpose-explanations

- Goal of purpose-explanations is to provide information about the function of a fact or some object.
- Example:
  - “What is the valve for?” – “The valve is used to seal the intake and exhaust ports.”



## Cognitive explanations

- Cognitive explanations explain the activities of the system.
- Examples:
  - Action explanations:  
“Why was this seat post selected?” – “For the given price, only one other seat post was available. But this was too short.”
  - Negative explanations:  
“Why was no carrier chosen?” – “A carrier is only available for touring bikes. The user did not choose a touring bike.”



# Overview

---

- 1 Motivation
- 2 Kinds of Explanations
- 3 Explanation Goals

# Explanation goals

---

- Explain Why the Answer is a Good Answer (Justification)
- Explain How the System Reached the Answer (Transparency)
- Explain Why a Question Asked is Relevant (Relevance)
- Teach the User About the Domain (Learning)

## Illustrative Example (1)

- Scenario: Diagnostic system for engine failures; mixed initiative system
- Starting point: The system asks the user to fill in the specific gravity of the fuel in the tank, but the user does not know the concept.
- User: “What is the specific gravity?”

System: “It is a measure of the density of a liquid relative to the density of water, with water having a specific gravity of 1.0.”

Addressing a learning goal with a conceptual explanation (definition)

## Illustrative Example (2)

- User: “Why is the density of the fuel interesting?”

System: “Air/Fuel ratio is based on pounds of AIR to pounds of FUEL. Changing the specific gravity of your fuel can have a profound affect on the A/F ratio of your engine since carburetors are metered by volume.”

Addressing the relevance goal with a why-explanation

## Illustrative example (3)

- User: "How is this adjustment related to the problem of too much CO<sub>2</sub> in the exhaust gases?"

System: "Different fuels and engine setups require different spark lead times. When using a wrong timing, the combustion process is suboptimal, and this will lead to high CO<sub>2</sub> values." **Addressing the justification goal with a theoretical conceptual explanation**

## Illustrative example (4)

- User: "How did you come up with this answer?"

System: "Looking at my case base we can see that the most relevant cases had problems with the spark lead time as well, with only the exception of those cases where the problem can be explained by the wrong type of fuel."

**Addressing the transparency goal with a how-explanation**

## Future work

---

- Generalizations / transfer from expert systems research results
- Reasoning knowledge vs. explanatory knowledge
- Contextual knowledge and personalization
- Implementation and evaluation