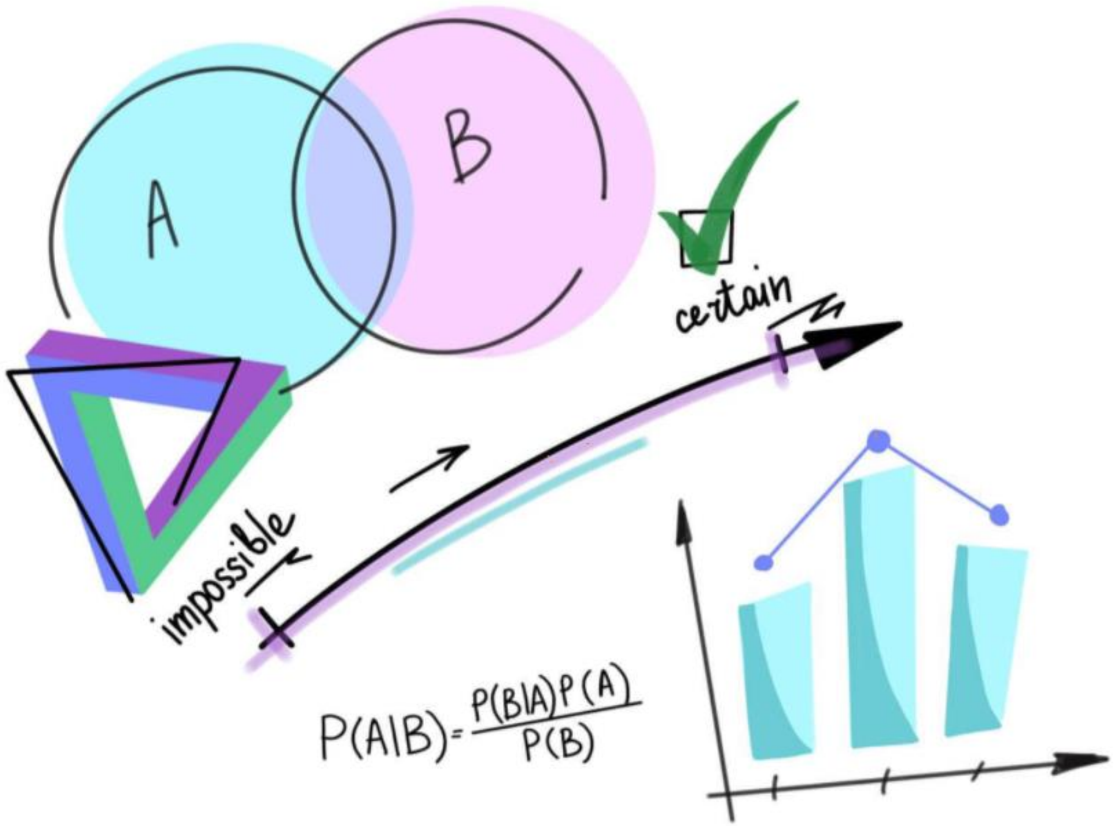
20IS603 Architecture of Intelligent Systems

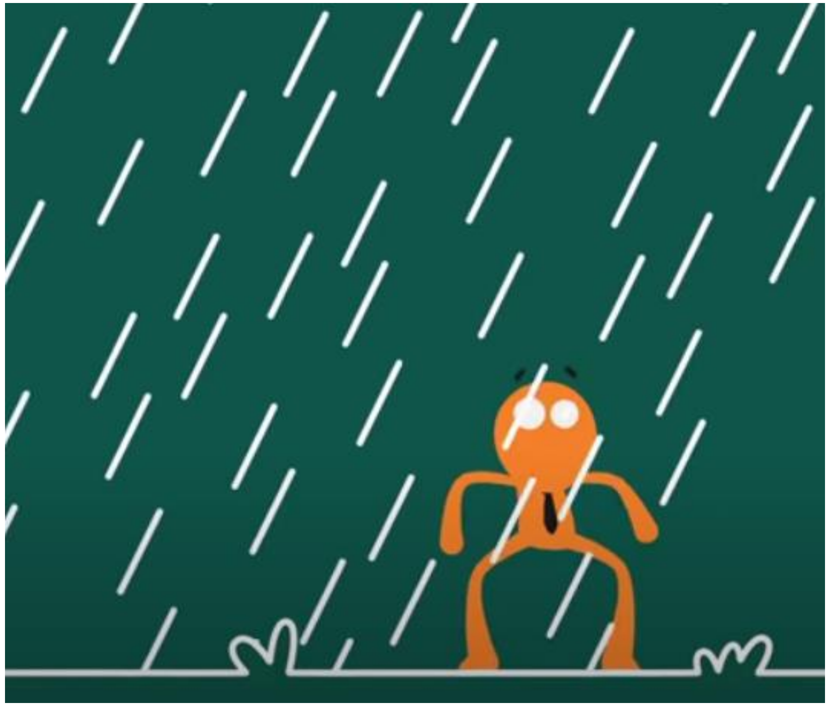


Handling Uncertainty in Expert Systems

Sources of uncertainty

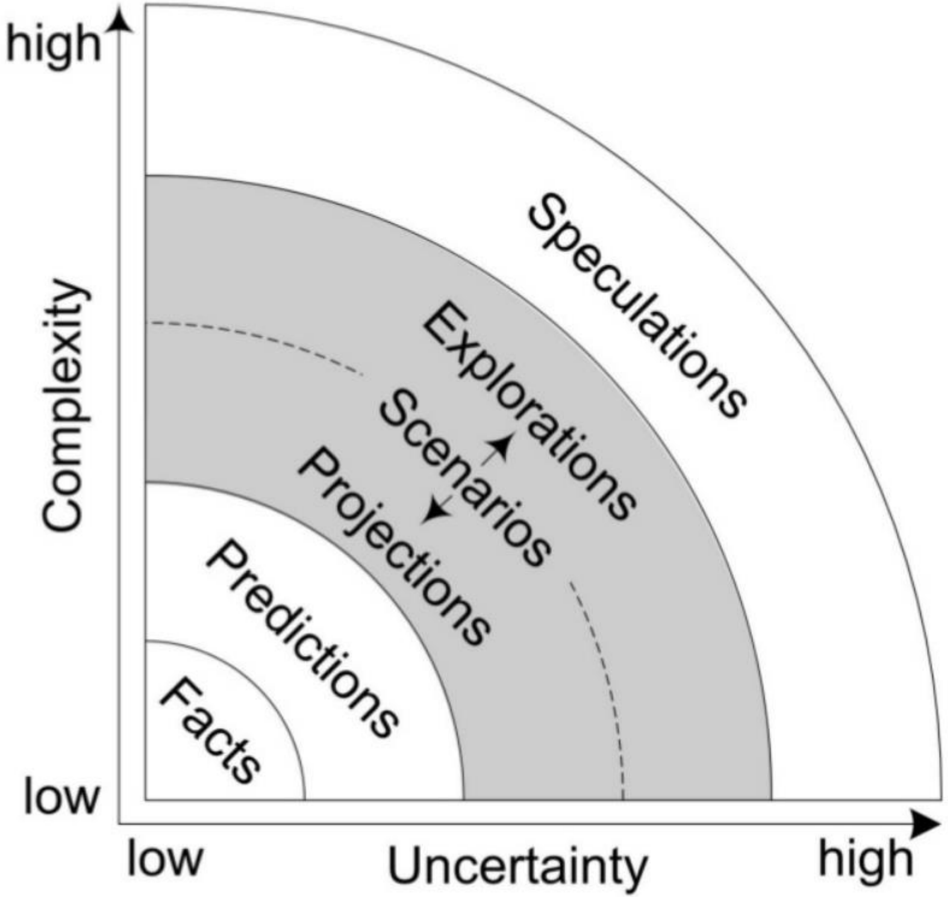
Bayesian Updating

Lecture #5



 Or  ?





high
low
Complexity
low
Uncertainty
high
Facts
Predictions
Projections
Scenarios
Explorations
Speculations

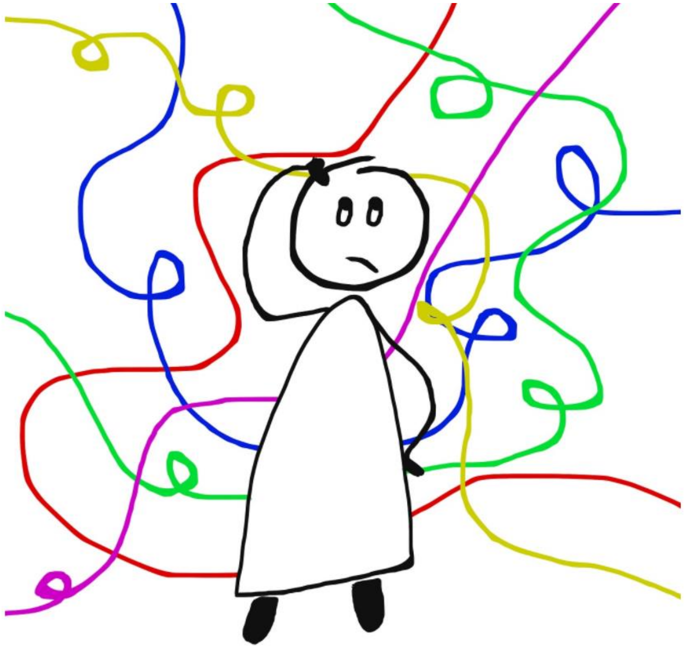
Uncertainty

- Lack of information to formulate a decision
- Result in making poor or bad decisions
- Dealing with uncertainty requires reasoning under uncertainty
- Deductive reasoning – deals with exact facts and exact conclusions
- Inductive reasoning – not as strong as deductive – premises support the conclusion but do not guarantee it.
- Number of methods to pick the best solution in light of uncertainty.
- When dealing with uncertainty, have to settle for just a good solution.

Sources of Uncertainty

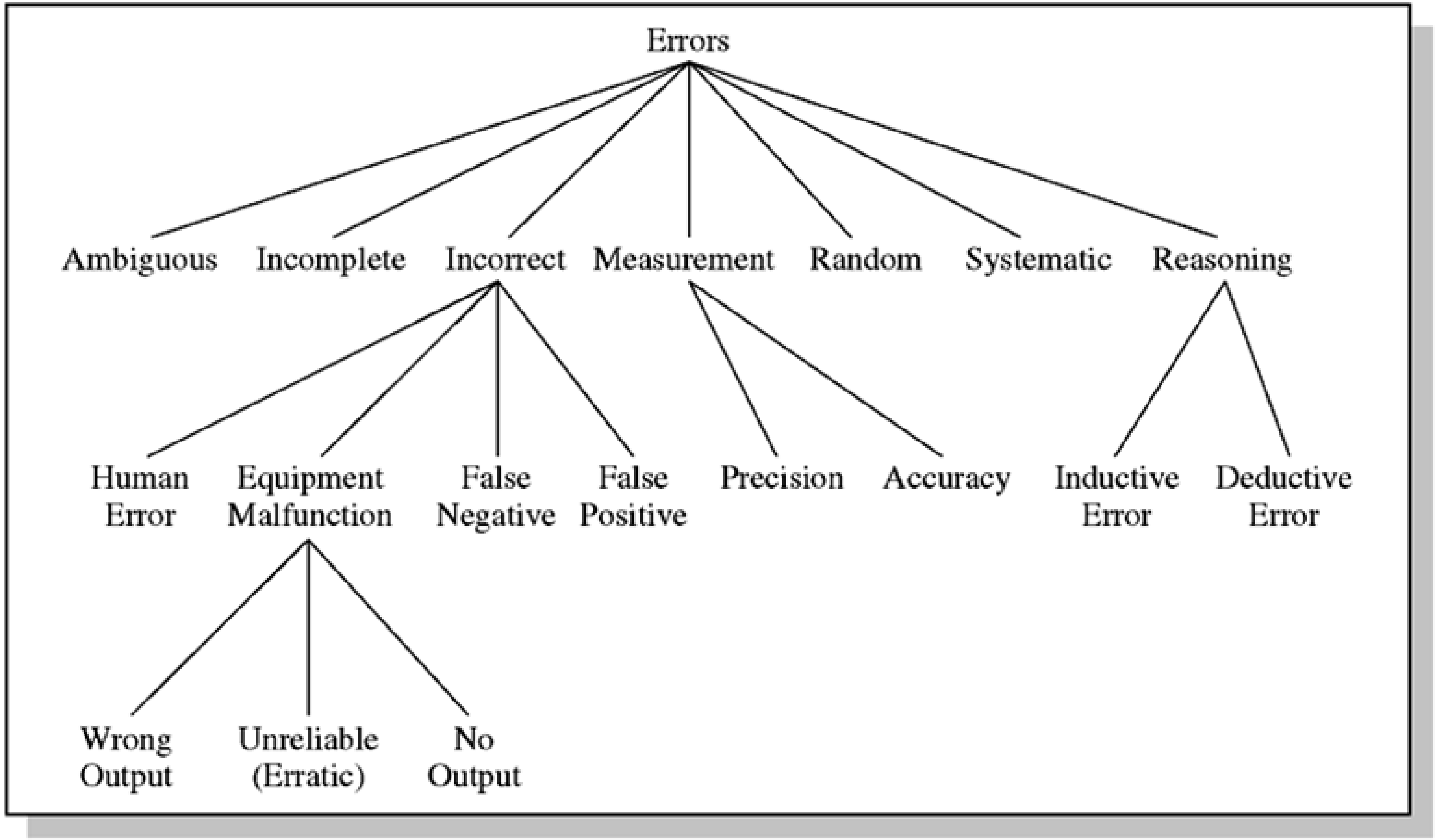
- Weak implications
  - Correlations between IF (condition) and THEN (action) parts of the rules - Vague associations
- Imprecise language
  - Different terms are used with the same meaning or a term has multiple (different) meanings – imprecise data
  - Interpreted in more than one way - ambiguity
- Unknown data
  - Information is not sufficient or missing to make a decision
- Combination of different expert views
  - Multiple experts have contradictory opinions and produce conflicting rules

Forms of Uncertainty



- Uncertainty in the rule itself
  - rules based on heuristics will be uncertain
- Uncertainty in the evidence
  - the evidence may come from a source that is not totally reliable
  - Inductive arguments can never be proven correct
- Use of vague language
  - imprecision in the representation language

Types of errors contributing to uncertainty



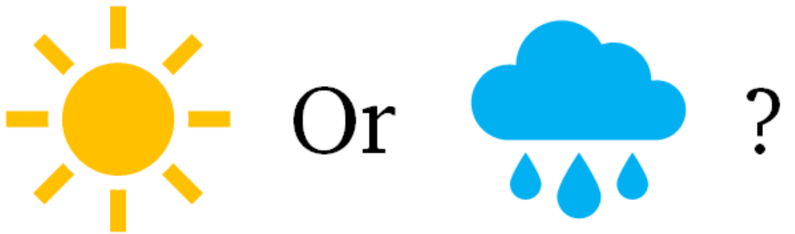
Types of errors contributing to uncertainty

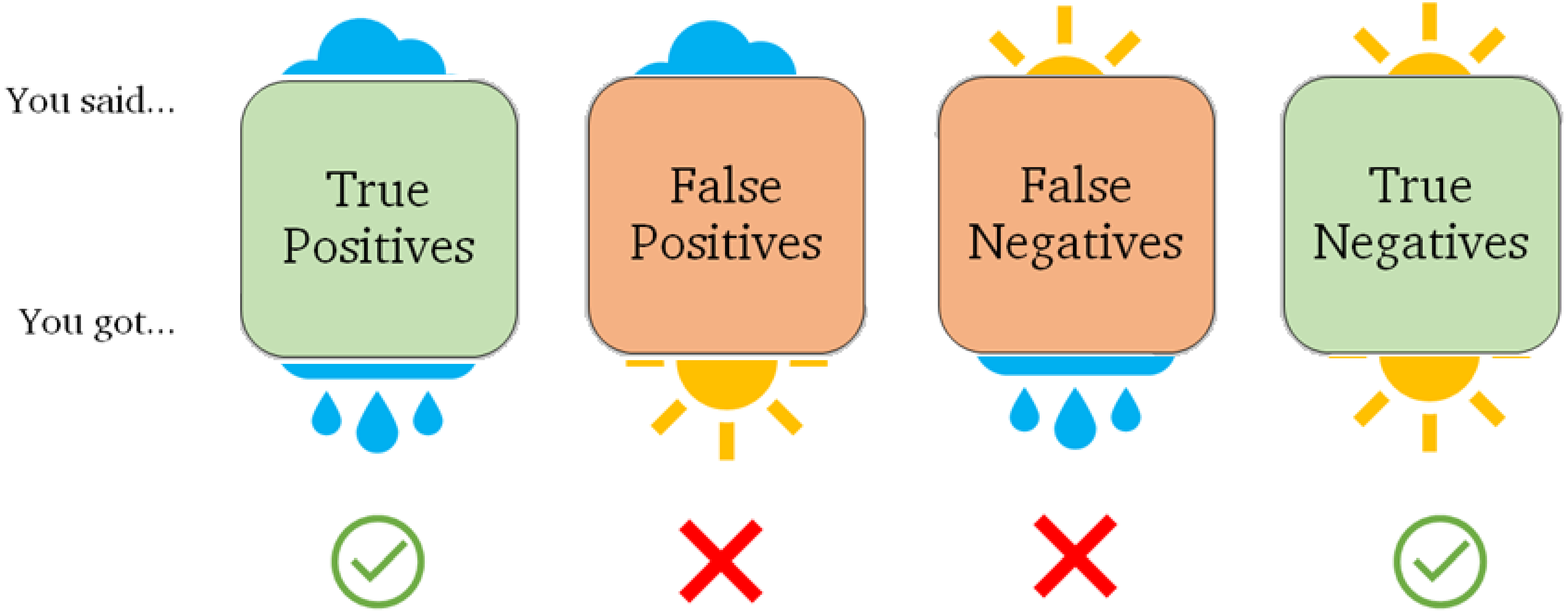
Examples of Common Types of Errors

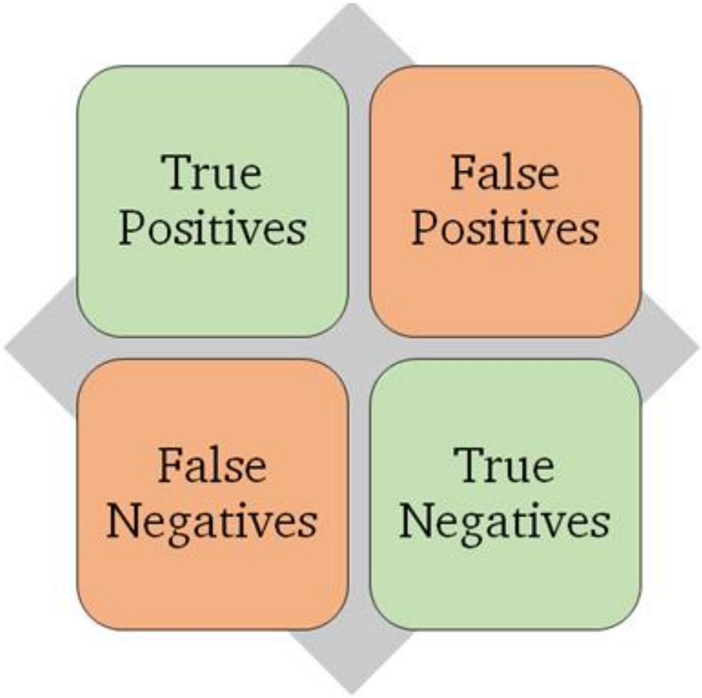
Example	Error	Reason
Turn the valve off	Ambiguous	What valve?
Turn valve-1	Incomplete	Which way?
Turn valve-1 off	Incorrect	Correct is on
Valve is stuck	False positive	Valve is not stuck
Valve is not stuck	False negative	Valve is stuck
Turn valve-1 to 5	Imprecise	Correct is 5.4
Turn valve-1 to 5.4	Inaccurate	Correct is 9.2
Turn valve-1 to 5.4 or 6 or 0	Unreliable	Equipment error
Valve-1 setting is 5.4 or 5.5 or 5.1	Random Error	Statistical Fluctuation
Valve-1 setting is 7.5	Systematic Error	Miscalibration
Valve-1 is not stuck because it has never been stuck before	Invalid Induction	Valve is stuck
Output is normal and so valve-1 is in good condition	Invalid deduction	Valve is stuck in open position

Types of errors contributing to uncertainty

False positive and False negative







Classical Probability Theory

Three axioms of probability

Axiom 1: $0 \leq P(E) \leq 1$ for any event E

Axiom 2: $\sum P(E) = 1$

Sum of all events that do not affect each other – mutually exclusive events is 1

Axiom 3: $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

The probability of the union of mutually exclusive events is the sum of the probabilities of the individual events.

Classical Probability Theory

Three axioms of probability

Axiom 1: $0 \leq P(E) \leq 1$ for any event E

Axiom 2: $\sum P(E) = 1$

Sum of all events that do not affect each other – mutually exclusive events is 1

Axiom 3: $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

The probability of the union of mutually exclusive events is the sum of the probabilities of the individual events.

For pairwise independent events:

$$P(A \cap B) = P(A)\,P(B)$$

Additive law:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

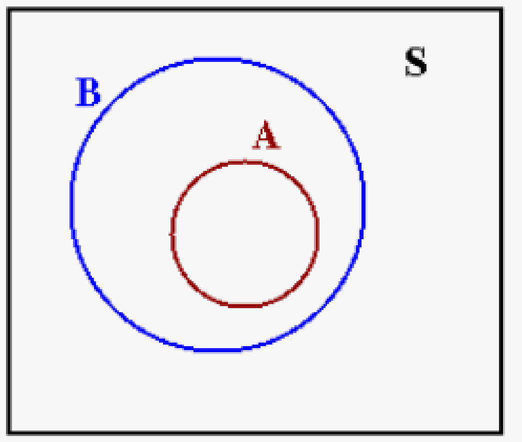
A and B are two events which are not mutually independent

Conditional Probability

- Compute the probabilities of events under certain conditions
- Conditional probability of A given that B has occurred

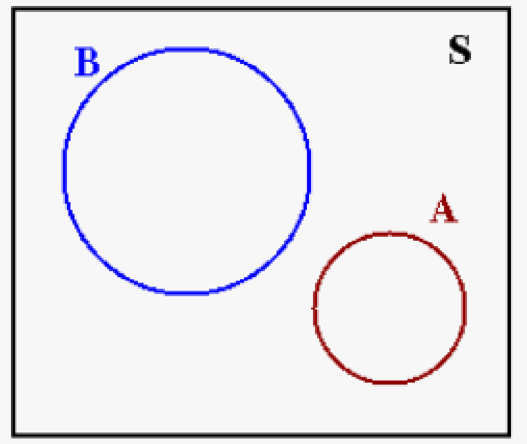
$$P(A|B)$$

- States the probability that A is true, given that we already know that A is true



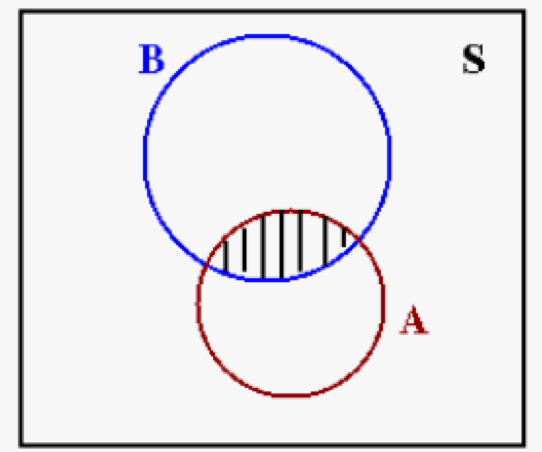
A subset of B

P(A|B) = 1



A and B are disjoint

P(A|B) = 0



Sets A and B intersect

P(A|B) = P(A and B)/P(B)

Conditional Probability

- Conditional probability of event A occurring given that event B has occurred

$$\mathrm{P}(A|B) = \frac{\text{the number of times } A \text{ and } B \text{ can occur}}{\text{the number of times } B \text{ can occur}}$$

- The number of times A and B can occur, or the probability that both A and B will occur, is called the joint probability of A and B.
- The number of ways B can occur is the probability of B, P(B)

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\textit{Joint probability of A and B}}{\textit{Probability of B}}$$

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(B \cap A)}{\mathrm{P}(A)}$$

Joint probability is commutative, $\mathrm{P}(A \cap B) = \mathrm{P}(B \cap A)$

Bayesian Reasoning

Bayes' Theorem for Conditional Probabilities

H: *Hypothesis*

E: *Evidence*

$$P(H \mid E) = \frac{P(E \mid H)\,P(H)}{P(E)}$$

T. Bayes.

Bayesian Rule

- The probability of an event based on prior knowledge of the conditions that might be relevant to the event

$$P(A \cap B) = P(B \cap A)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(A \cap B) = P(A|B) \times P(B)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \qquad P(B \cap A) = P(B|A) \times P(A)$$

product rule

$$\boxed{P(A|B) = \frac{P(B|A) * P(A)}{P(B)}}$$

Bayesian Rule

- The probability of an event based on prior knowledge of the conditions that might be relevant to the event

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

P(A) - prior probability of A

"prior" that it does not take into account any information about B.

P(A|B) - posterior probability

derived from or depends upon the specified value of B

P(B|A) - Likelihood

the chance that something will happen

P(B) – prior or marginal likelihood of B

acts as a normalizing constant

Bayesian Rule

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \qquad \text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}}$$

Given that A is true, B must either be true or false

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \qquad P(\neg B|A) = \frac{P(A|\neg B) \cdot P(\neg B)}{P(A)}$$

Then, P(B|A) + P(~B|A) = 1

$$1 = \frac{P(A|B) \cdot P(B)}{P(A)} + \frac{P(A|\neg B) \cdot P(B)}{P(A)}$$

$$\therefore P(A) = P(A|B) \cdot P(B) + P(A|\neg B) \cdot P(\neg B)$$

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\neg B) \cdot P(\neg B)}$$

Bayesian Rule

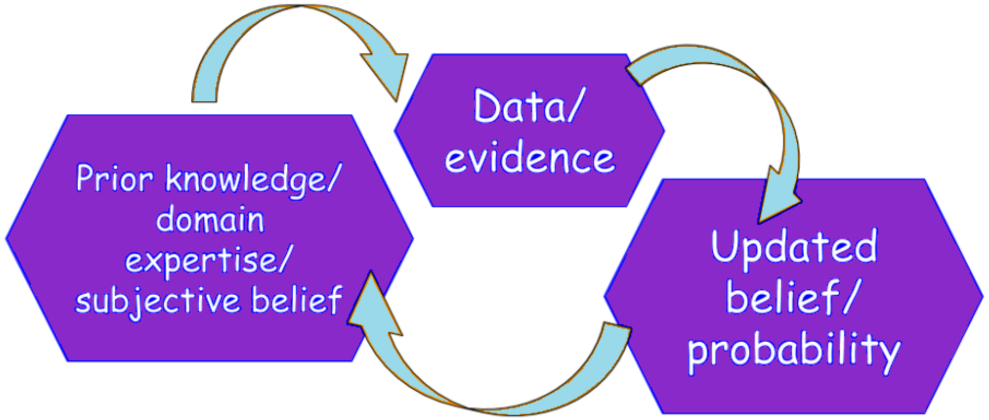
$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E|H) \times P(H) + P(E|\sim H) \times P(\sim H)}$$

The probability of ~H is given by $P(\sim H) = 1 - P(H)$

- P(H), the current probability of the hypothesis. If this is the first update for this hypothesis, then P(H) is the prior probability.
- P(E|H), the conditional probability that the evidence is present, given that the hypothesis is true.
- P(E|~H), the conditional probability that the evidence is present, given that the hypothesis is false.

Bayesian Updating

- Technique for updating probability in the light of evidence for or against the hypothesis
- Updating is cumulative - if the probability of a hypothesis has been updated in the light of one piece of evidence, the new probability can then be updated further by a second piece of evidence.
- The evidence is a symptom, and the hypothesis is a diagnosis.
- P(H), P(E|H), and P(E|~H) values are needed for all the different hypotheses and evidence covered by the rules.
- Performs abduction (i.e., determining causes) using deductive information (i.e., the likelihood of symptoms, effects, or evidence).



Bayesian Updating

Likelihood Ratio

- The odds O(H) of a given hypothesis H are related to its probability P(H) by

$$\boxed{O(H) = \frac{P(H)}{P(\sim H)} = \frac{P(H)}{1 - P(H)}}$$

$$P(H) = \frac{O(H)}{O(H) + 1}$$

- An assertion that is absolutely certain (having a probability of 1) has infinite odds
- O(H|E), the conditional odds of H given E, is

$$O(H|E) = \frac{P(H|E)}{P(\sim H|E)}$$

Bayesian Updating

Likelihood Ratio

$$P(H|E)=\frac{P(H)\times P(E|H)}{P(E)} \qquad P(\sim H|E)=\frac{P(\sim H)\times P(E|\sim H)}{P(E)}$$

dividing above equations

$$\frac{P(H|E)}{P(\sim H|E)}=\frac{P(H)\times P(E|H)}{P(\sim H)\times P(E|\sim H)}$$

We know that,

$$O(H|E)=\frac{P(H|E)}{P(\sim H|E)}$$

Substituting this,

$$O(H|E)=\frac{P(H)\times P(E|H)}{P(\sim H)\times P(E|\sim H)}$$

$$\boxed{O(H|E)=A\times O(H)}$$

where $A=\frac{P(E|H)}{P(E|\sim H)}$ - *affirms* weight of evidence E

Likelihood of sufficiency

Bayesian Updating

Likelihood Ratio

Absence of evidence - *denies* weight *D* of evidence E

$$\boxed{\mathrm{O(H|{\sim}E)} = D \times \mathrm{O(H)}}$$

where $$D = \frac{\mathrm{P({\sim}E|H)}}{\mathrm{P({\sim}E|{\sim}H)}} = \frac{1 - \mathrm{P(E|H)}}{1 - \mathrm{P(E|{\sim}H)}}$$ Likelihood of necessity

$$\boxed{\mathrm{P}(H|E) = \frac{O(H|E)}{1 + O(H|E)}}$$

If a given piece of evidence E has an affirms weight A that is greater than 1, then its denies weight must be less than 1 and vice versa:

A>1 implies D<1
A<1 implies D>1

If A<1 and D>1, then the absence of evidence is supportive of a hypothesis.

Bayesian Updating

Dealing with Uncertain Evidence

- Evidence is either definitely present (i.e., has a probability of 1) or definitely absent (i.e., has a probability of 0).
- If the probability of the evidence lies between these extremes, then the confidence in the conclusion must be scaled appropriately
- Reasons for the evidence to be uncertain:
  - The evidence could be an assertion generated by another uncertain rule, and which therefore has a probability associated with it.
  - The evidence may be in the form of data that are not totally reliable, such as the output from a sensor.
- Assume that E was asserted by another rule whose evidence was B, where B is certain (has probability 1).
- Given the evidence B, the probability of E is P(E|B).

$$P(H|B) = P(H|E) \times P(E|B) + P(H|\sim E) \times [1 - P(E|B)]$$

useful only if Bayes' theorem is being used directly

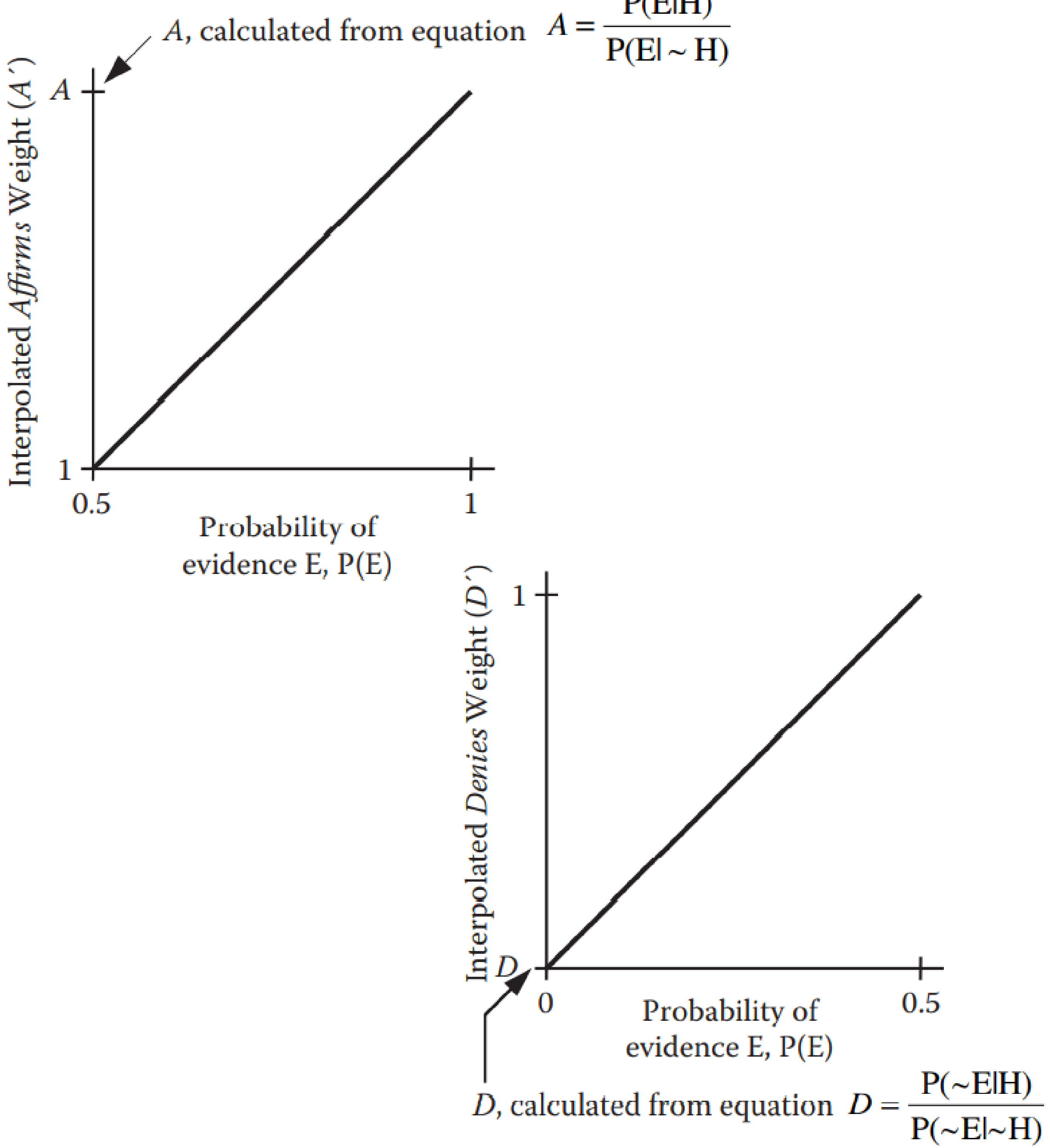
Bayesian Updating

Dealing with Uncertain Evidence

- When using likelihood ratios, one technique is to modify the *affirms* and *denies* weights to reflect the uncertainty in E.
- Achieved by interpolating the weights linearly as the probability of E varies between 1 and 0.
- While P(E) is greater than 0.5, the affirms weight is used, and when P(E) is less than 0.5, the denies weight is used.
- Over the range of values for P(E), A′ and D′ vary between 1 (neutral weighting) and A and D, respectively.

$$A' = [2(A - 1) \times P(E)] + 2 - A$$

$$D' = [2(1 - D) \times P(E)] + D$$



Bayesian Updating

Combining Evidence

- Bayesian updating centered on combining several pieces of evidence that support the same hypothesis
- If n pieces of evidence are found that support a hypothesis H, then the formal restatement of the updating equation is:

$$O(H|E_1\&E_2\&E_3 \ldots E_n) = A \times O(H)$$

$$\text{where} \quad A = \frac{P(E_1\&E_2\&E_3\ldots E_n|H)}{P(E_1\&E_2\&E_3\ldots E_n|\sim H)}$$

- Usefulness of this pair of equations is doubtful since pieces of evidence which will be available to support the hypothesis H is not known.
- Expressions for A covering all possible pieces of evidence Ei is written, as well as all combinations of the pairs $E_i\&E_j$, of the triples $E_i\&E_j\&E_k$, of quadruples $E_i\&E_j\&E_k\&E_m$, and so on - unrealistic requirement when the number of possible pieces of evidence is large

Bayesian Updating

Combining Evidence

- The problem becomes much more manageable if it is assumed that all pieces of evidence are statistically independent.
- This assumption of Bayesian updating in knowledge-based systems is rarely accurate.
- Statistical independence of two pieces of evidence (E_1 and E_2) means that the probability of observing E_1 given that E_2 has been observed is identical to the probability of observing E_1 given no information about E_2
- Statistical independence of E1 and E2 is defined as

$$P(E_1|E_2) = P(E_1) \quad \text{and} \quad P(E_2|E_1) = P(E_2)$$

- If n pieces of evidence are found that support or oppose H, then the updating equations are

$$O(H|E_1\&E_2\&E_3 \ldots E_n) = A_1 \times A_2 \times A_3 \times \ldots \times A_n \times O(H)$$

$$O(H|\sim E_1\&\sim E_2\&\sim E_3 \ldots \sim E_n) = D_1 \times D_2 \times D_3 \times \ldots \times D_n \times O(H)$$

Bayesian Updating

Combining Evidence

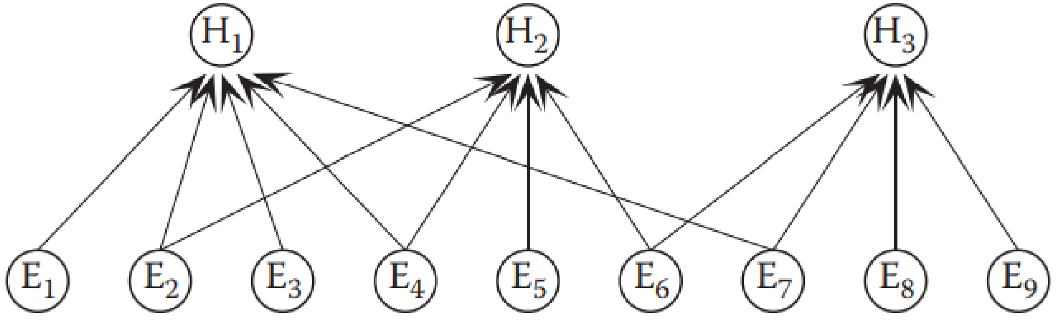
- Problems arising from the interdependence of pieces of evidence can be avoided if the rule base is properly structured.
- Pieces of evidence known to be dependent on each other, should not be combined in a single rule.
- Instead, assertions—and the rules that generate them—should be arranged in a hierarchy from low-level input data to high-level conclusions, with many levels of hypotheses between - does not limit the amount of evidence considered in reaching a conclusion, but controls the interactions between the pieces of evidence.
- Inference networks are a convenient means of representing the levels of assertions from input data through intermediate deductions to final conclusions.

Bayesian Updating

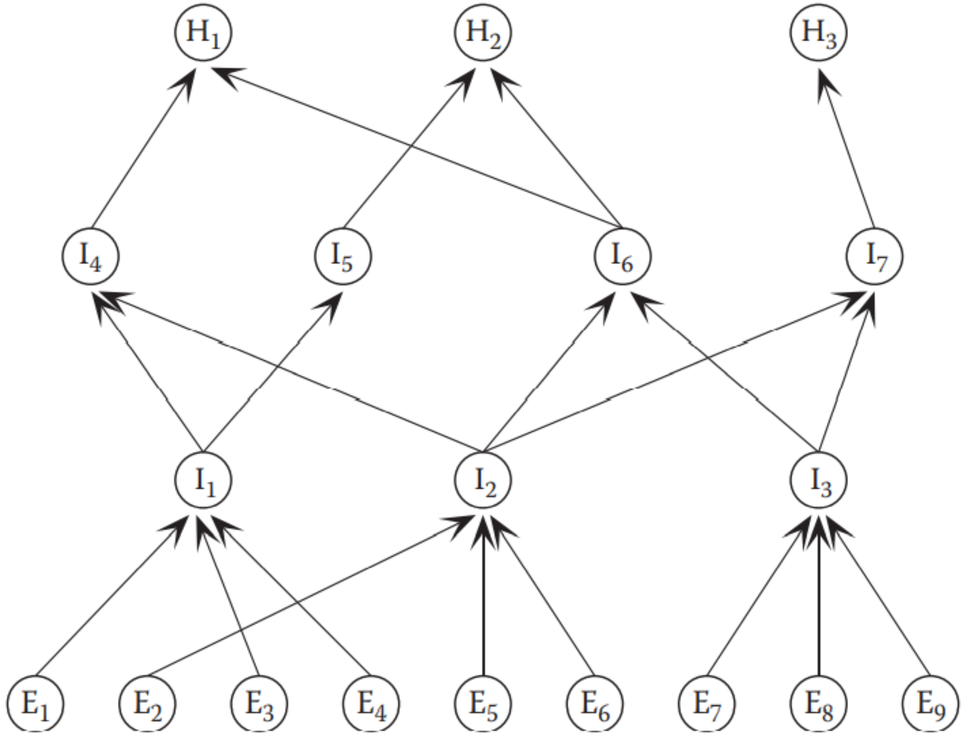
Combining Evidence

- All the evidence that is relevant to particular conclusions are drawn together in a single rule for each conclusion - produces a shallow network - only be reliable if there was no dependence between the input data

- Deeper network with intermediate levels between input data and conclusions



A shallow Bayesian inference network (E_i = evidence, H_i = hypothesis)



A deeper Bayesian inference network (I_i = intermediate hypothesis)

Bayesian Rules with Production Rules

- Practical rule-based system - mix uncertain rules with production rules
- If a production rule contains multiple pieces of evidence that are independent from each other, their combined probability can be derived from standard probability theory
- Consider, a rule with conjoined pieces of independent evidence:

  if <evidence E_1> and <evidence E_2> then <hypothesis H_3>

  Probability of hypothesis H_3 is given by

$$P(H_3) = P(E_1) \times P(E_2)$$

- Production rules containing independent evidence that is disjoined

  if <evidence E_1> or <evidence E_2> then <hypothesis H_3>

  Probability of hypothesis H_3 is given by

$$P(H_3) = P(E_1) + P(E_2) - (P(E_1) \times P(E_2))$$

Advantages of Bayesian Updating

- Based upon a proven statistical theorem.
- Likelihood is expressed as a probability (or odds)
- Requires deductive probabilities, which are generally easier to estimate than abductive ones. The user supplies values for the probability of evidence (the symptoms) given a hypothesis (the cause), rather than the reverse.
- Likelihood ratios and prior probabilities can be replaced by sensible guesses.
- Evidence for and against a hypothesis (or the presence and absence of evidence) can be combined in a single rule by using affirms and denies weights.
- Linear interpolation of the likelihood ratios can be used to take account of any uncertainty in the evidence.
- The probability of a hypothesis can be updated in response to more than one piece of evidence.

Disadvantages of Bayesian Updating

- Prior probability of an assertion must be known or guessed at.
- Conditional probabilities must be measured or estimated or, failing those, guesses must be taken at suitable likelihood ratios.
- The single probability value for the truth of an assertion tells us nothing about its precision.
- The addition of a new rule that asserts a new hypothesis often requires alterations to the prior probabilities and weightings of several other rules.
- The assumption that pieces of evidence are independent is often unfounded. The only alternatives are to calculate affirms and denies weights for all possible combinations of dependent evidence, or to restructure the rule base so as to minimize these interactions.
- The linear interpolation technique for dealing with uncertain evidence is not mathematically justified
- Representations based on odds, as required to make use of likelihood ratios, cannot handle absolute truth, that is, odds = ∞

Example #1

- Control of a power station boiler

```
rule r3_1a
  if release_valve is stuck
  then task becomes clean_release_valve.

rule r3_2a
  if warning_light is on
  then release_valve becomes stuck.

rule r3_3a
  if pressure is high
  then release_valve becomes stuck.

rule r3_4a
  if temperature is high
  and water_level is not low
  then pressure becomes high.
```

List out the hypothesis H and evidence E.

Example #1

- Probability values

H	E	P(H)	P(E|H)	P(E|~H)	O(H)	*A*	*D*
release valve needs cleaning	release valve is stuck	—	—	—	—	—	—
release valve is stuck	warning light is on	0.02	0.88	0.4	0.02	2.20	0.20
release valve is stuck	pressure is high	0.02	0.85	0.01	0.02	85.0	0.15
pressure is high	temperature is high	0.1	0.90	0.05	0.11	18.0	0.11
pressure is high	water level is low	0.1	0.05	0.5	0.11	0.10	1.90

Example #1

- New rules

```
uncertainty_rule r3_1b
  if release_valve is stuck
  then task becomes clean_release_valve.

uncertainty_rule r3_2b
  if warning_light is on (affirms 2.20; denies 0.20)
  then release_valve becomes stuck.

uncertainty_rule r3_3b
  if pressure is high (affirms 85.0; denies 0.15)
  then release_valve becomes stuck.

uncertainty_rule r3_4b
  if temperature is high (affirms 18.0; denies 0.11)
  and water_level is low (affirms 0.10; denies 1.90)
  then pressure becomes high.
```

Example #1

- set of input data

```
water_level is not low.
warning light is on.
temperature is high.
```

assume that the rules fire in the following order:

$$r3\_4b \rightarrow r3\_3b \rightarrow r3\_2b \rightarrow r3\_1b$$

Example #1

```
uncertainty_rule r3_4b
H = pressure is high;                          O(H) = 0.11
E1 = temperature is high;                      A1 = 18.0
E2 = water_level is low;                       D2 = 1.90
O(H|(E1&~E2)) = O(H) × A1 × D2 = 3.76
/* Updated odds of "pressure is high" are 3.76 */

uncertainty_rule r3_3b
H = release_valve is stuck;                    O(H) = 0.02
E = pressure is high;                          A = 85.0
Because E is not certain (O(E) = 3.76, P(E) = 0.79), the
inference engine must calculate an interpolated value A'
for the affirms weight of E (see Section 3.2.5).
A'= [2(A-1) × P(E)] + 2 - A = 49.7
O(H|(E)) = O(H) × A' = 0.99
/* Updated odds of "release_valve is stuck" are 0.99, */
/* corresponding to a probability of approximately 0.5
*/

uncertainty_rule r3_2b
H = release_valve is stuck;                    O(H) = 0.99
E = warning_light is on;                       A = 2.20
O(H|(E)) = O(H) × A = 2.18
/* Updated odds of "release_valve is stuck" are 2.18 */

uncertainty_rule r3_1b
H = task is clean_release_valve
E = release_valve is stuck;
O(E)= 2.18 implies O(H)= 2.18
/* This is a production rule, so the conclusion is
asserted with the same probability as the evidence. */
/* Updated odds of "task is clean_release_valve" are
2.18 */
```

Final Conclusion

O(task is clean_release_valve) = 2.18

or

P(task is clean_release_valve) = 0.69