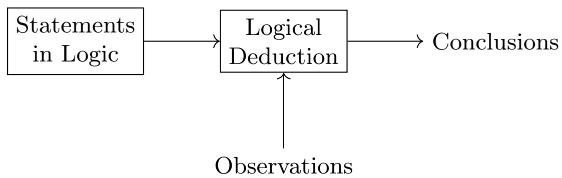
In the beginning of AI, there was a debate between symbolic and numeric representations.
Symbolic representation is now standard, though numeric representation is used in Bayesian Networks.

Once knowledge representation was decided, acquisition needed to be decided.
We can model knowledge, whether procedurally or declaratively.
We can learn knowledge from information in the world.
The former approach is often used in AI, whereas the latter is used in ML.
This is changing however; progress in modern day requires more broad knowledge.

We will be discussing the **classical model of logic**



Consider the example problem of the WumpusWorld:

1. The grid contains a goal, pits, and a Wumpus
2. Cells adjacent to pits are breezy
3. Cells adjacent to the Wumpus smell
4. The Wumpus and pits kill you

P			
Breeze	OK		
A	Stench	W	

The logic depends on two components:

1. Prior Knowledge (Rule/Problem Structure)
2. Observation (Learning/Breeze and Stench)

This is a canonical **logical** approach; ≡ if something is deduced, it MUST be true.
It is more complicated than many animals can do!

Propositional Logic

There are two major components to propositional logic:

1. **Syntax** (≡ Grammar)
2. **Semantics** (≡ Meaning)

Even if the syntax doesn't match, semantics can tell if two statements are equal

Syntax

A syntax consists of three elements:

1. boolean variables $(X_1, ..., X_N)$
2. logical connectives
   (a) $\wedge \equiv$ AND
   (b) $\vee \equiv$ OR
   (c) $\neg \equiv$ NOT
   (d) $\implies \equiv$ IMPLIES
   (e) $\iff \equiv$ IFF
3. Sentences
   (a) if S is a sentence, then $\neg S$ is a sentence.
   (b) if $S_1 \& S_2$ are sentences, then each of the following are compound sentences:
      - $S_1 \wedge S_2$
      - $S_1 \vee S_2$
      - $S_1 \implies S_2$
      - $S_1 \iff S_2$

We use a few terms to refer to elements of a syntax:

- $X, \neg X$ are called **literals**
  - X is the **positive literal**

- $\neg X$ is the **negative literal**

- $X \wedge Y$ is called a **conjunction**
  - X, Y are called the **conjuncts**
- $X \vee Y$ is called a **disjunction**
  - X, Y are called the **disjuncts**
- $X \implies Y$ is called a **premise**
  - X is called the **premise antecedent**
  - Y is called the **implicant**
- $\neg Y \implies X$ is called the **contrapositive** (in relation to the premise)

We have a few logical rules to work with sentences:

1. $a \implies B = \neg a \vee B$
2. $\neg(a \wedge B) = \neg a \vee \neg B$
3. $\neg(a \vee B) = \neg a \wedge \neg B$

Let's try to capture the following state with logic:

	B_1	
B_2	P	B_3
	B_4	

Consider the example problem of the WumpusWorld:
We can represent this in (at least) two ways:

1. $P \implies (B_1 \vee B_2 \vee B_3 \vee B_4)$
2. $P \implies (B1 \wedge B2 \wedge B3 \wedge B4)$

The latter is **stronger**, since it is more specific.
The former is true, but doesn't model all we know.
This is a common problem when writing AI.

Normal Forms

Normal forms can be categorized in one of two ways:

1. a grammar is **practical/restricted** if there are logics it cannot represent.
2. a grammar is **universal** if it can represent any theoretical logic.

The normal forms we will cover are:

1. **Conjunctive Normal Form***
   $[(A \wedge \neg B) \vee (A \wedge \neg D \wedge E) \vee ...]$
   We call a disjunction of literals a **clause**.
   Therefore, this form is a **conjunction of clauses**.
2. **Disjunctive Normal Form***
   $[(A \vee \neg B) \wedge (A \vee \neg D \vee E) \vee ...]$
   We call a conjunction of literals a **term**.
   Therefore, this form is a **disjunction of terms**.
3. **Horn Form**
   $[(A \vee \neg B \vee \neg C) \vee (\neg A \vee \neg B \vee \neg C)]$
   A **horn clause** is a clause with at most 1 positive literal.
   A clause $A \vee \neg B \vee \neg C$ can be written as $B \wedge C \implies A$
   $B \wedge C$ is called the **body**; A is called the **head**.
4. **Negation Normal Form***
   $[(A \wedge \neg D \wedge E) \vee (A \vee \neg C \vee D) \wedge (A \wedge \neg B)]$
   Allows either clauses or terms, but only allows ($\neg$) on variables, not sentences.
   Are often represented as circuits.
   Is tractable iff all entries to all AND gates share no variables.
   This subset of NNFs are called **Decomposable (DNNF)**.

(* is used to represent grammars that are universal)
Fun fact: SAT on a CNF is hard but SAT on a DNF is easy.
Horn Form makes SAT linear, therefore Horn Form is **tractable**.

We can do these SAT problems in two ways:

1. **Forward Chaining** (data driven) $\equiv$ iterate over the clauses; if the body is true, infer the head
2. **Backward Chaining** (Goal Driven) $\equiv$ recursively attempt to prove the goal to the head's body

In the ideal case, backward chaining is much faster than even linear!

Semantics

Semantics are concerned with the notion of **worlds**.
$\equiv$ a set containing a fixed truth value for all variables/preposition symbols
We would like to know if a sentence holds in a given world.
If a holds in world w, we say $w \models a$ (equivalent to "world w contains a").

Consider the following example:

	E	A	B
W_1	T	T	T
W_2	T	T	F
W_3	T	F	T
W_4	T	F	F
W_5	F	T	T
W_6	F	T	F
W_7	F	F	T
W_8	F	F	F

Which worlds do the following hold in?

- E? $\{1, 2, 3, 4\}$
- $\neg E$? $\{5, 6, 7, 8\}$
- $\neg B$? $\{3, 4, 7, 8\}$
- $\neg B \wedge \neg E$? $\{7, 8\}$
- A? $\{1, 3, 5, 7\}$
- $((\neg E \wedge \neg B) \vee A)$? $\{1, 3, 5, 7, 8\}$
- $((E \vee B) \implies A)$? $\{1, 3, 5, 7, 8\}$

Therefore,

- $W_1 \vee W_3 \vee W_5 \vee W_7 \vee W_8 \models (E \vee B) \implies A$
- $W_2 \vee W_4 \vee W_6 \not\models (E \vee B) \implies A$

We can thus derive that for sets of worlds S and T, if $S \models a$ && $T \models b$, then

1. $S \cap T \models (a \wedge b)$
2. $S \cup T \models (a \vee b)$
3. $S/T \models (a \wedge \neg b)$

If we thus assert the property $(a \wedge b)$, we can prune the set of worlds down to $S \cap T$.
If we get to one world, we know all; if we get to none, we have an inconsistency.
We can use these properties to build a pretty effective SAT Solver.

We denote the **meaning/model** of $\alpha M(\alpha)$ and define it by $M(\alpha) = \{w | w \models \alpha\}$.
Thus we should be able to translate the following easily:

1. α is equivalent to $\beta \rightarrow M(\alpha) = M(\beta)$
2. α is contradictory $\rightarrow M(\alpha) = \{\}$
3. α is a **tautology**/is valid $\rightarrow M(\alpha) = \{w\}$
4. α & β are mutually exclusive $\rightarrow M(\alpha) \cap M(\beta) = \{\}$
5. α implies β ($\alpha \implies \beta$) $\rightarrow M(\alpha) \subset M(\beta)$