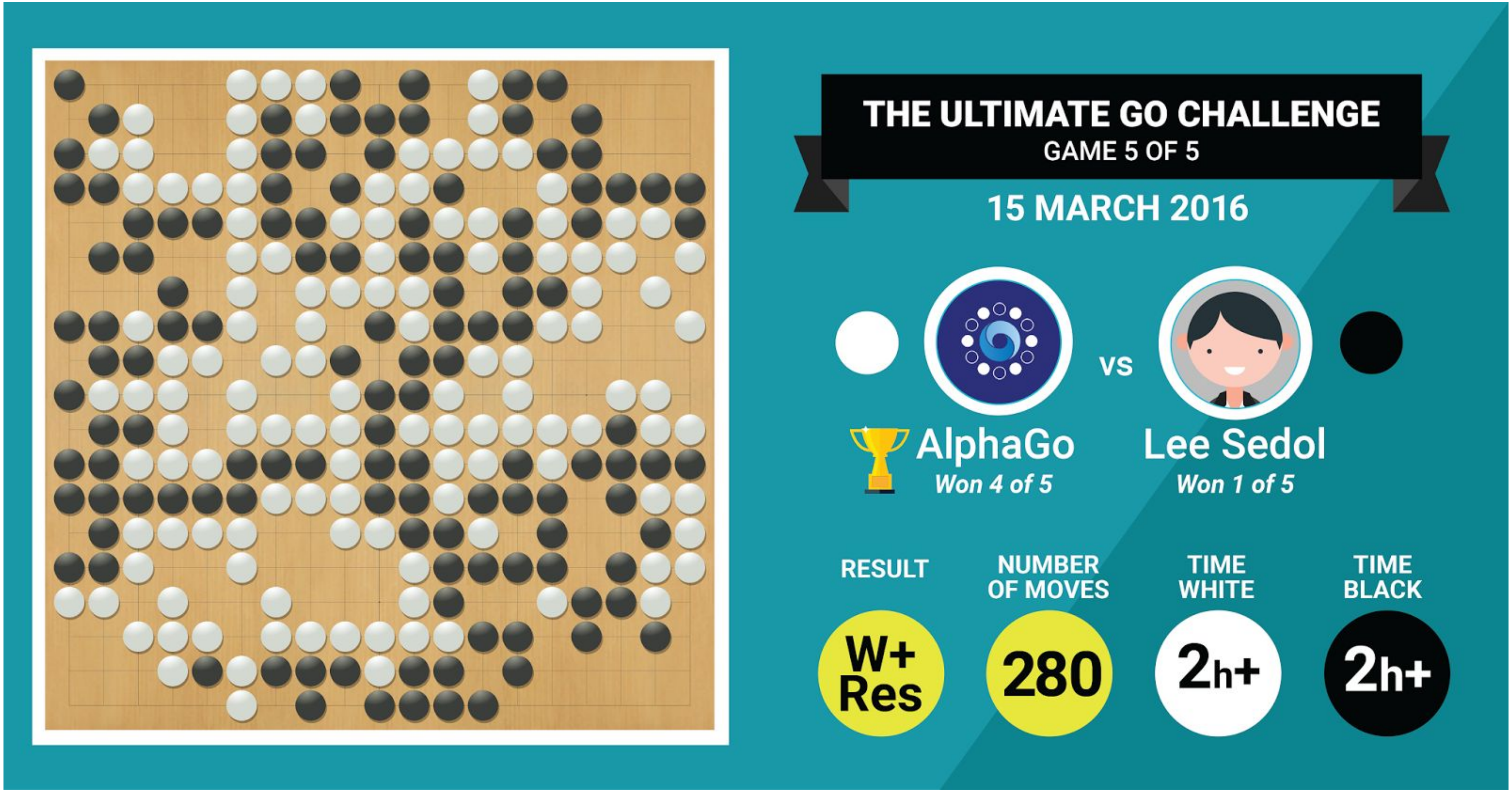
Logical Agents

Update: http://googleasiapacific.blogspot.co.uk/



Review

Logical agents apply inference to a knowledge base to derive new information and make decisions

Basic concepts of logic:

- syntax: formal structure of sentences
- semantics: truth of sentences with respect to models
- entailment: necessary truth of one sentence given another
- inference: deriving sentences from other sentences

Wumpus World

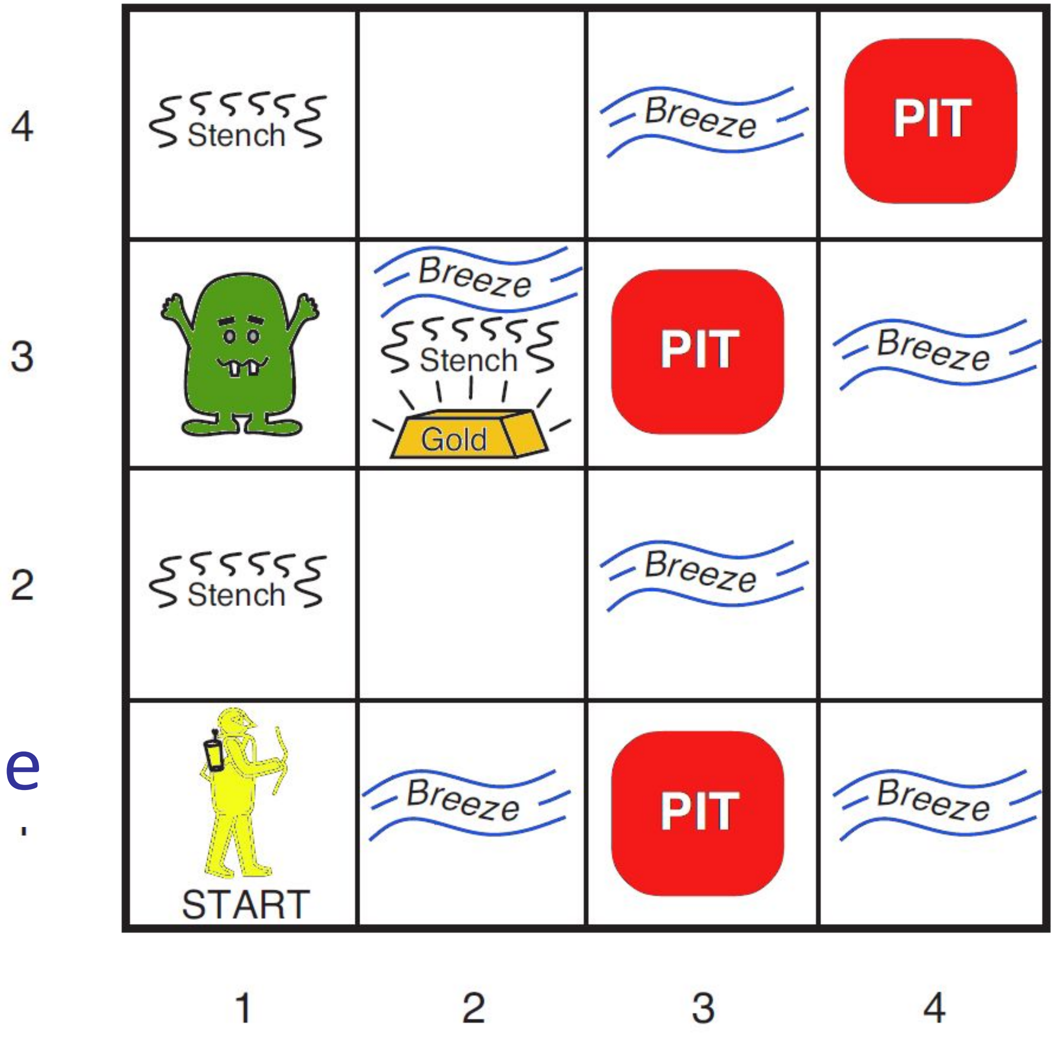
A cave with rooms connected by passageways.

The wumpus is somewhere in the cave and eats anyone who enters it room.

The wumpus doesn't move.

Some rooms contain bottomless pits.

There is a heap of gold somewhere in the cave.



Applying Inference

Legitimate (sound) generation of new sentences from old

Proof = a sequence of inference rule applications

Typically require transformation of sentences into a normal form

Inference Rules

Whenever any sentences of the form $\alpha \Rightarrow \beta$ and α are given, then β can be inferred. (Modus Ponens).

$$\frac{\alpha \Rightarrow \beta, \alpha}{\beta}$$

KB:

$B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})$

$B_{1,1}$

$(P_{1,2} \vee P_{2,1})$

Inference Rules

Any of the sentences can be inferred from a conjunction of sentences (and elimination):

$$\frac{\alpha \land \beta}{\alpha}$$

It is sunny and I have an umbrella

It is sunny

Inference Rules - Unit Resolution

Whenever sentences of the form $\alpha \vee \beta$ and $\neg\alpha$ are given, then β can be inferred.

$$\frac{\alpha \vee \beta,\ \neg\alpha}{\beta}$$

There is a pit in [1, 2] or [2, 1]: $P_{1,2} \vee P_{2,1}$

There is no pit in [1, 2]: $\neg P_{1,2}$

There is a pit in [2, 1]: $P_{2,1}$

Inference Rules - Resolution

$$\frac{A \lor B,\ \neg A \lor C}{B \lor C}$$

Suppose A is false.

Apply unit resolution: $A \lor B$, $\neg A$ entail B. So B is true

Suppose A is true, then $\neg A$ is false.

Apply unit resolution: $\neg A \lor C$ and $\neg(\neg A)$ entail C. So C is true

Since A must be either true or false, either B or C is true. $B \lor C$ is true

Inference Rules - Resolution

$$\frac{A \vee B,\ \neg A \vee C}{B \vee C}$$

We take two sentences of the form ($A \vee B$ and $\neg A \vee C$)

We produce a new sentence containing the literals of the original sentences except for the two complementary literals (A and $\neg A$):

$B \vee C$

Inference Rules - Resolution

We take two sentences. We produce a new sentence containing the literals of the original clauses except for the two complementary literals.

Which of the sentences below is inferred when we apply resolution to: A V B V C and ¬B V C V D:

A. A V B V C
B. A V B V C V D
C. A V C V D
D. B V C V D
E. A V B V D

Resolution Algorithm

To show that $KB \models \alpha$, we show $KB \land \neg\alpha$ is unsatisfiable.

1. Convert $KB \land \neg\alpha$ to CNF (Conjunctive Normal Form)
2. Apply resolution rule repeatedly
3. At the end: empty clause - unsatisfiable ($A \land \neg A$)

Resolution is sound and complete for propositional logic

Resolution in the Wumpus World

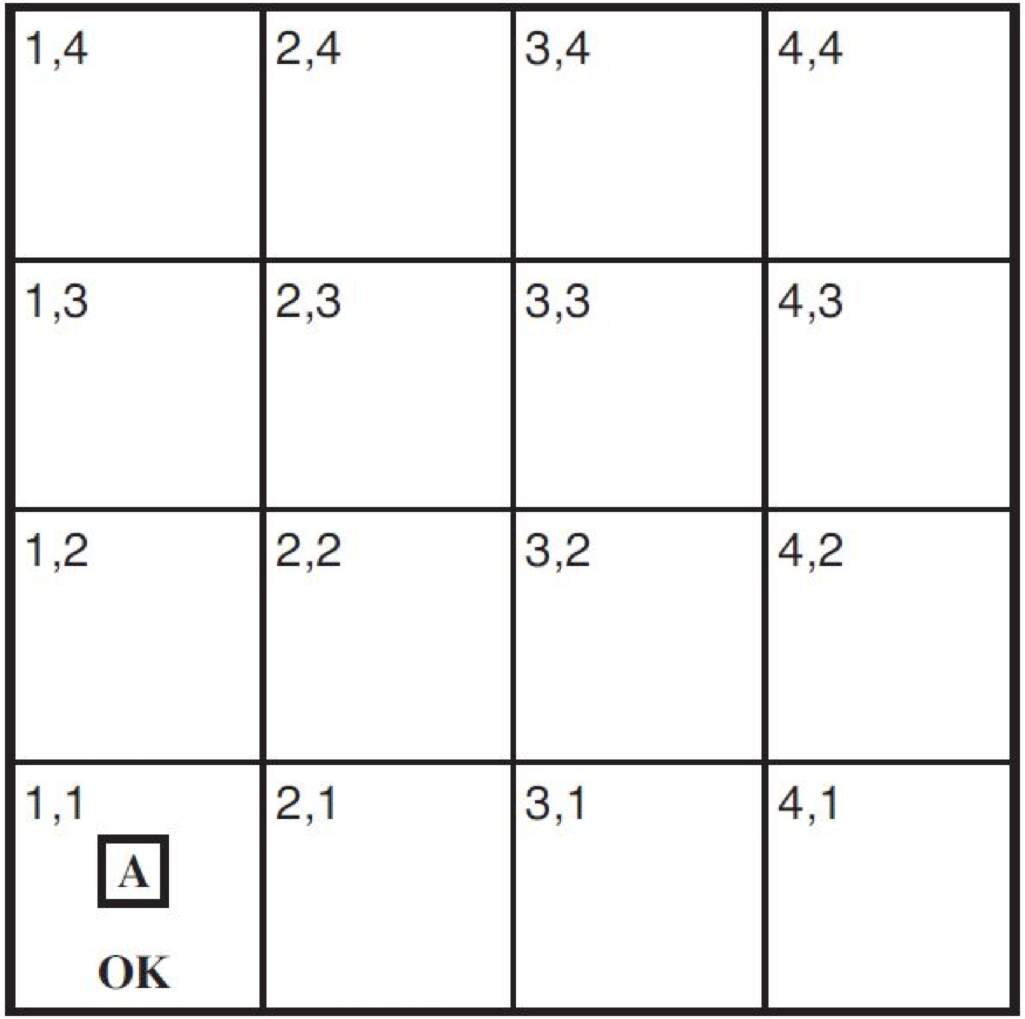
A room is breezy if and only if there is an adjacent pit:

$B_{1,1} \Leftrightarrow \quad (P_{1,2} \vee P_{2,1})$

Percept: $\neg B_{1,1}$

Our goal is to show that:

$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1}) \wedge \neg B_{1,1}$ entail $\neg P_{1,2}$



Resolution Algorithm

To show that *KB* $\models \alpha$, we show KB $\wedge \neg\alpha$ is unsatisfiable.

To show that $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1}) \wedge \neg B_{1,1} \models \neg P_{1,2}$

we show that $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1}) \wedge \neg B_{1,1} \wedge P_{1,2}$ is unsatisfiable

1. Convert KB to CNF
2. Apply resolution rule repeatedly to KB(in CNF) $\wedge \neg\alpha$
3. At the end: we get an empty clause - unsatisfiable $(P_{1,2} \wedge \neg P_{1,2})$

Step 1: Conversion to CNF

$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1}) \wedge \neg B_{1,1}$

1. Eliminate $\Leftrightarrow$, replacing $\alpha \Leftrightarrow \beta$ with $(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)$.

   $(B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1}) \wedge \neg B_{1,1}$

2. Eliminate $\Rightarrow$, replacing $\alpha \Rightarrow \beta$ with $\neg\alpha \vee \beta$.

   $(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg(P_{1,2} \vee P_{2,1}) \vee B_{1,1}) \wedge \neg B_{1,1}$

3. Move $\neg$ inwards using de Morgan's rules:

   $(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge ((\neg P_{1,2} \wedge \neg P_{2,1}) \vee B_{1,1}) \wedge \neg B_{1,1}$

4. Apply distributivity law ($\vee$ over $\wedge$) and flatten:

   $(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}) \wedge \neg B_{1,1}$

Step 2: Apply Resolution

Apply resolution rule repeatedly to KB(in CNF) $\wedge$ $\neg\alpha$

KB = $(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}) \wedge \neg B_{1,1}$

$\alpha = \neg P_{1,2}$

Apply resolution rule repeatedly to:

$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}) \wedge \neg B_{1,1} \wedge P_{1,2}$

Apply Resolution - Resolve

$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}) \wedge \neg B_{1,1} \wedge P_{1,2}$

$\boxed{\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}}$ $\boxed{\neg P_{1,2} \vee B_{1,1}}$ $\boxed{\neg P_{2,1} \vee B_{1,1}}$ $\boxed{\neg B_{1,1}}$ $\boxed{P_{1,2}}$

Apply Resolution

$(\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land (\neg P_{1,2} \lor B_{1,1}) \land (\neg P_{2,1} \lor B_{1,1}) \land \neg B_{1,1} \land P_{1,2}$

$\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}$	$\neg P_{1,2} \lor B_{1,1}$	$\neg P_{2,1} \lor B_{1,1}$	$\neg B_{1,1}$	$P_{1,2}$

Apply Resolution

$(\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land (\neg P_{1,2} \lor B_{1,1}) \land (\neg P_{2,1} \lor B_{1,1}) \land \neg B_{1,1} \land P_{1,2}$

$\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}$	$\neg P_{1,2} \lor B_{1,1}$	$\neg P_{2,1} \lor B_{1,1}$	$\neg B_{1,1}$	$P_{1,2}$

Apply Resolution

$$(\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land (\neg P_{1,2} \lor B_{1,1}) \land (\neg P_{2,1} \lor B_{1,1}) \land \neg B_{1,1} \land P_{1,2}$$

$\boxed{\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}}$ $\boxed{\neg P_{1,2} \lor B_{1,1}}$ $\boxed{\neg P_{2,1} \lor B_{1,1}}$ $\boxed{\neg B_{1,1}}$ $\boxed{P_{1,2}}$

$\boxed{P_{1,2} \lor P_{2,1} \lor \neg P_{1,2}}$

Apply Resolution

$$(\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land (\neg P_{1,2} \lor B_{1,1}) \land (\neg P_{2,1} \lor B_{1,1}) \land \neg B_{1,1} \land P_{1,2}$$

$\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}$ | $\neg P_{1,2} \lor B_{1,1}$ | $\neg P_{2,1} \lor B_{1,1}$ | $\neg B_{1,1}$ | $P_{1,2}$

$P_{1,2} \lor P_{2,1} \lor \neg P_{1,2}$

Apply Resolution

$$(\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land (\neg P_{1,2} \lor B_{1,1}) \land (\neg P_{2,1} \lor B_{1,1}) \land \neg B_{1,1} \land P_{1,2}$$

$\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}$

$\neg P_{1,2} \lor B_{1,1}$

$\neg P_{2,1} \lor B_{1,1}$

$\neg B_{1,1}$

$P_{1,2}$

$P_{1,2} \lor P_{2,1} \lor \neg P_{1,2}$

$\neg B_{1,1} \lor P_{2,1} \lor B_{1,1}$

Apply Resolution

$$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}) \wedge \neg B_{1,1} \wedge P_{1,2}$$

$\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}$ | $\neg P_{1,2} \vee B_{1,1}$ | $\neg P_{2,1} \vee B_{1,1}$ | $\neg B_{1,1}$ | $P_{1,2}$

$P_{1,2} \vee P_{2,1} \vee \neg P_{1,2}$ | $\neg B_{1,1} \vee P_{2,1} \vee B_{1,1}$

Apply Resolution

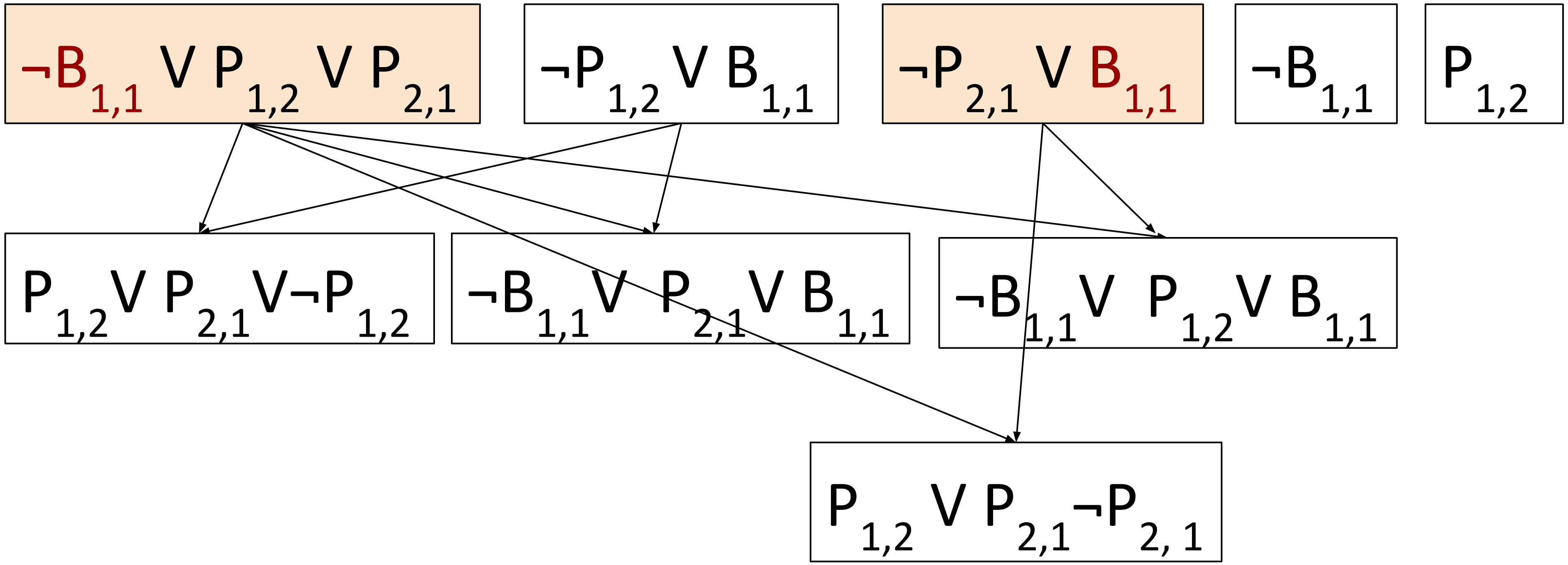
$$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}) \wedge \neg B_{1,1} \wedge P_{1,2}$$

$\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}$	$\neg P_{1,2} \vee B_{1,1}$	$\neg P_{2,1} \vee B_{1,1}$	$\neg B_{1,1}$	$P_{1,2}$

$P_{1,2} \vee P_{2,1} \vee \neg P_{1,2}$	$\neg B_{1,1} \vee P_{2,1} \vee B_{1,1}$	$\neg B_{1,1} \vee P_{1,2} \vee B_{1,1}$

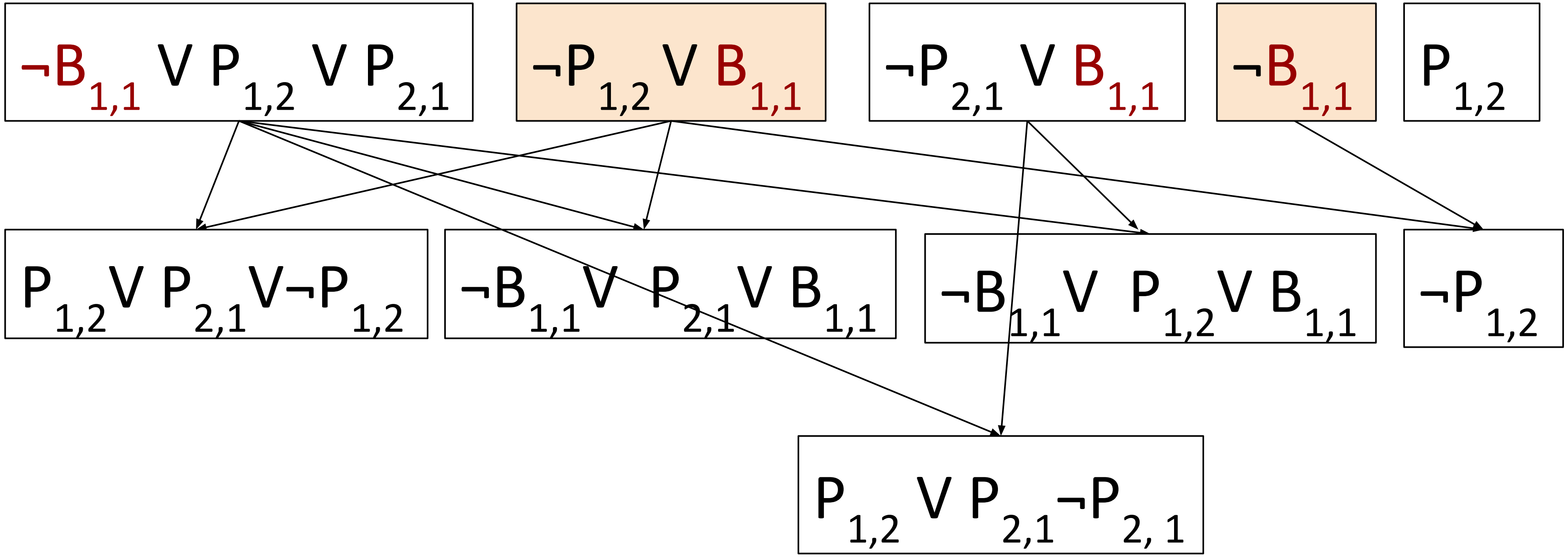
Apply Resolution

$$(\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land (\neg P_{1,2} \lor B_{1,1}) \land (\neg P_{2,1} \lor B_{1,1}) \land \neg B_{1,1} \land P_{1,2}$$



Apply Resolution

$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}) \wedge \neg B_{1,1} \wedge P_{1,2}$



Apply Resolution

$$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}) \wedge \neg B_{1,1} \wedge P_{1,2}$$

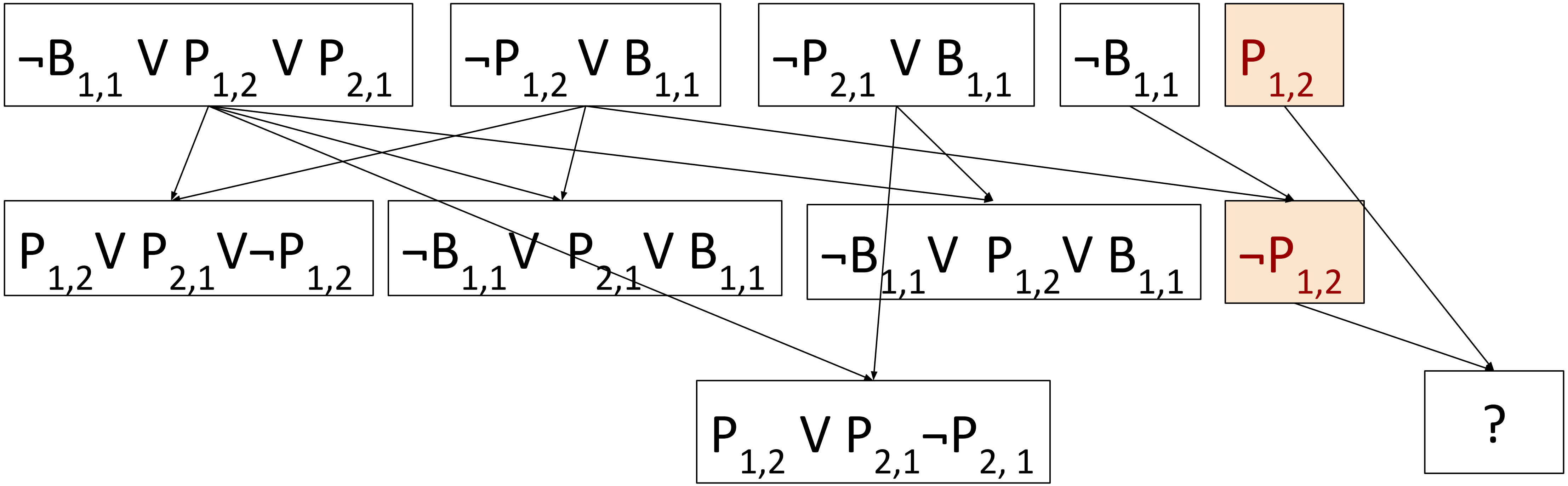
$\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}$ | $\neg P_{1,2} \vee B_{1,1}$ | $\neg P_{2,1} \vee B_{1,1}$ | $\neg B_{1,1}$ | $P_{1,2}$

$P_{1,2} \vee P_{2,1} \vee \neg P_{1,2}$ | $\neg B_{1,1} \vee P_{2,1} \vee B_{1,1}$ | $\neg B_{1,1} \vee P_{1,2} \vee B_{1,1}$ | $\neg P_{1,2}$

$P_{1,2} \vee P_{2,1} \neg P_{2,1}$

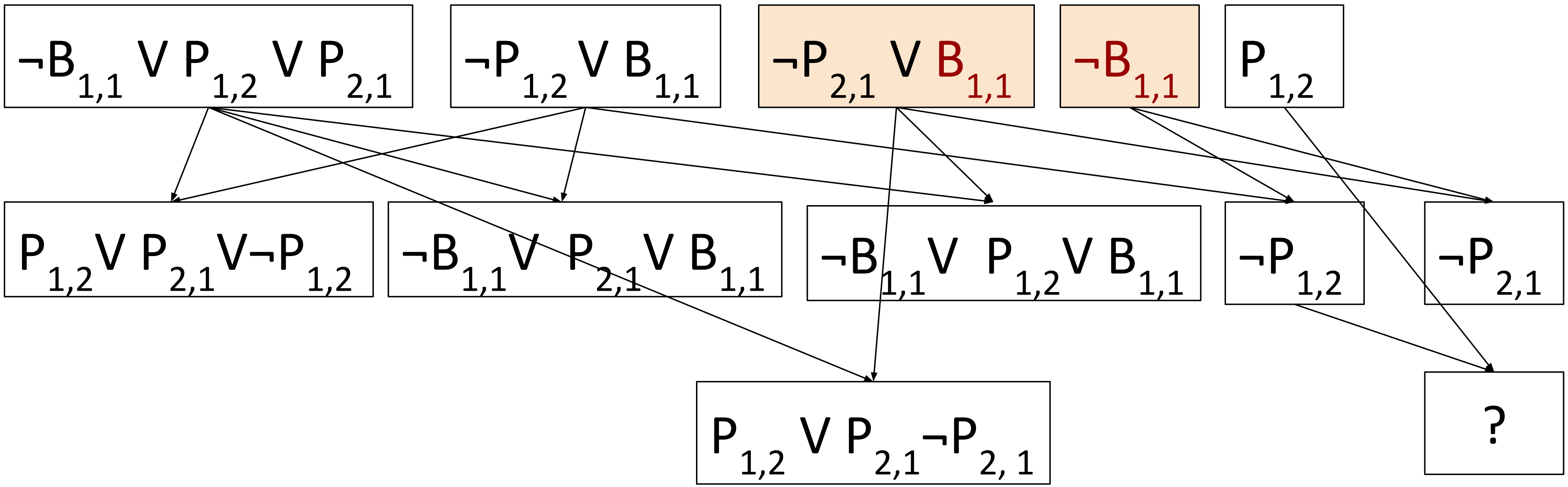
Apply Resolution

$$(\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land (\neg P_{1,2} \lor B_{1,1}) \land (\neg P_{2,1} \lor B_{1,1}) \land \neg B_{1,1} \land P_{1,2}$$



Apply Resolution

Note that we can also generate:



Limitations

Propositional language: powerful language

Supports efficient inference mechanisms for determining validity and satisfiability

Lacks expressive power

Represents events that are true or false in our world - no uncertainty

Cannot represent objects and properties or relations between objects

Cannot represent 'pits cause breezes in adjacent rooms' except by writing one sentence for each room

First-Order Logic

Propositional logic assumes world contains facts

First-order logic (like natural language) assumes the world contains:

- **Objects**: people, numbers, courses, rooms...
- **Relations**: (between objects, or describing properties) Located(cs156, dh450), Adjacent([1, 1], [1, 2]), Breezy([1, 2])
- **Functions**: (mapping from object to object):

  BestFriend(...), Instructor(..)

Syntax of First-Order Logic

Constants represent objects:

[1, 1], 2, SJSU, Anna, Bob, Ryan, Joe, ...

Predicates represent relations:

Adjacent([1, 1], [1, 2]), Sibling(Ryan, Joe), Dating(Anna, Bob)

predicate($term_1$,...,$term_n$) is true
iff the objects referred to by $term_1$,...,$term_n$
are in the relation referred to by the predicate

Functions: represent a mapping between objects:

sqrt, BestFriend

Syntax of First-Order Logic

Variables also represent objects:

x, y, a, b, student, room, . . .

Connectives (same as in propositional logic)

$\land \;\lor\; \neg \Rightarrow \Leftrightarrow$

Equality: =

refer to the same object

Quantifiers:

universal: $\forall$

existential: $\exists$

Syntax of First-Order Logic

Term:

- constant: Anna, [1, 1], [1, 2]
- variable: x, student, room
- function($term_1$, . . . ,$term_n$): BestFriend(Anna)

Atomic sentence:

- predicate($term_1$, . . . ,$term_n$):
  Adjacent([1, 1], [1, 2]), Dating(Ryan, BestFriend(Anna))
- term1 = term2: room = [1, 2], BestFriend(Anna) = Jenny

Syntax of First-Order Logic

Complex sentence:

- made from atomic sentences using connectives

  $\neg S$, $S1 \land S2$, $S1 \lor S2$, $S1 \Rightarrow S2$, $S1 \Leftrightarrow S2$

  Adjacent([1, 1], [1, 2]) $\land$ Breezy([1, 2])

  Dating(Ryan, BestFriend(Anna)) $\land$ BestFriend(Anna) = Jenny $\Rightarrow$ Dating(Ryan, Jenny)

Universal Quantification

$\forall$ *<variables>* *<sentence>*

Everyone at SJSU is smart:

$\forall x \; At(x, SJSU) \Rightarrow Smart(x)$

$\forall x$ P is true in a model *m* iff P is true with *x* being each possible object in the model

Equivalent to: At(Anna, SJSU) ⇒ Smart(Anna)

∧ At(Bob, SJSU) ⇒ Smart(Bob)

∧ At(Jenny, SJSU) ⇒ Smart(Jenny)

∧ At(Joe, SJSU) ⇒ Smart(Joe)

∧ At(Ryan, SJSU) ⇒ Smart(Ryan)

∧ ...

Universal Quantification

Typically $\Rightarrow$ is the main connective with $\forall$

Common mistake: using $\wedge$ as the main connective with $\forall$:

$$\forall x\ At(x, SJSU) \wedge Smart(x)$$

Everyone is at SJSU and everyone is smart

Existential Quantification

∃ *<variables>* *<sentence>*

Someone at SJSU is tall: $\exists x\ At(x, SJSU) \land Tall(x)$

$\exists x\ P$ is true in a model *m* iff P is true with *x* being some possible object in the model

Equivalent to:

$$
\begin{aligned}
& At(Anna, SJSU) \land Tall(Anna) \\
\lor\ & At(Bob, SJSU) \land Tall(Bob) \\
\lor\ & At(Jenny, SJSU) \land Tall(Jenny) \\
\lor\ & At(Joe, SJSU) \land Tall(Joe) \\
\lor\ & At(Ryan, SJSU) \land Tall(Ryan) \\
\lor\ & \ldots
\end{aligned}
$$

Existential Quantification

Typically, $\land$ is the main connective with $\exists$

Common mistake: using $\Rightarrow$ as the main connective with $\exists$:

$$\exists\, x \; \mathrm{At}(x, \mathrm{SJSU}) \Rightarrow \mathrm{Tall}(x)$$

is true if there is anyone who is not at SJSU

Properties of quantifiers

- $\forall x \, \forall y$ is the same as $\forall y \, \forall x$
- $\exists x \, \exists y$ is the same as $\exists y \, \exists x$
- $\exists x \, \forall y$ is not the same as $\forall y \, \exists x$

  $\exists x \, \forall y \, Loves(x,y)$

  There is a person who loves everyone in the world

  $\forall y \, \exists x \, Loves(x,y)$

  Everyone in the world is loved by at least one person

Back to our Wumpus World

Relations:

Pit(x): there is a pit in room x

Adjacent(x, y): rooms x and y are adjacent

Breezy(x): There is a breeze in room x

All rooms adjacent to a pit are breezy:

$$\forall r\ \forall s\ \mathrm{Pit}(r) \land \mathrm{Adjacent}(r, s) \Rightarrow \mathrm{Breezy}(s)$$

Any breezy room is adjacent to a pit

$$\forall s\ \mathrm{Breezy}(s) \Rightarrow \exists\ r\ \mathrm{Adjacent}(r,s) \land \mathrm{Pit}(r)$$