Knowledge

knowledge-based agents

agents that reason by operating on internal representations of knowledge

If it didn't rain, Harry visited Hagrid today.

Harry visited Hagrid or Dumbledore today, but not both.

Harry visited Dumbledore today.

Harry did not visit Hagrid today.

It rained today.

Logic

sentence

an assertion about the world
in a knowledge representation language

Propositional Logic

Proposition Symbols

P Q R

Logical Connectives

$\neg$
not

$\wedge$
and

$\vee$
or

$\rightarrow$
implication

$\leftrightarrow$
biconditional

Not ($\neg$)

P	$\neg P$
false	true
true	false

And ($\land$)

P	Q	$P \land Q$
false	false	false
false	true	false
true	false	false
true	true	true

Or ($\vee$)

P	Q	$P \vee Q$
false	false	false
false	true	true
true	false	true
true	true	true

Implication ($\rightarrow$)

P	Q	$P \rightarrow Q$
false	false	true
false	true	true
true	false	false
true	true	true

Biconditional ($\leftrightarrow$)

P	Q	$P \leftrightarrow Q$
false	false	true
false	true	false
true	false	false
true	true	true

model

assignment of a truth value to every propositional symbol (a "possible world") model

P: It is raining.

Q: It is a Tuesday.

$$\{P = \texttt{true}, Q = \texttt{false}\}$$

knowledge base

a set of sentences known by a knowledge-based agent

Entailment

$$\alpha \vDash \beta$$

In every model in which sentence α is true, sentence β is also true.

If it didn't rain, Harry visited Hagrid today.

Harry visited Hagrid or Dumbledore today, but not both.

Harry visited Dumbledore today.

Harry did not visit Hagrid today.

It rained today.

inference

the process of deriving new sentences from old ones

P: It is a Tuesday.

Q: It is raining.

R: Harry will go for a run.

KB: $(P \land \neg Q) \rightarrow R$ $\quad P \quad \neg Q$

Inference: R

Inference Algorithms

Does

$$\mathrm{KB} \models \alpha$$

?

Model Checking

Model Checking

- To determine if $\mathrm{KB} \vDash \alpha$:
  - Enumerate all possible models.
  - If in every model where KB is true, α is true, then KB entails α.
  - Otherwise, KB does not entail α.

P: It is a Tuesday. Q: It is raining. R: Harry will go for a run.

KB: $(P \land \neg Q) \rightarrow R$ $\quad P \quad \neg Q$

Query: R

P	Q	R	KB
false	false	false	
false	false	true	
false	true	false	
false	true	true	
true	false	false	
true	false	true	
true	true	false	
true	true	true	

P: It is a Tuesday. Q: It is raining. R: Harry will go for a run.

KB: $(P \land \neg Q) \rightarrow R$ P $\neg Q$

Query: R

P	Q	R	KB
false	false	false	false
false	false	true	false
false	true	false	false
false	true	true	false
true	false	false	false
true	false	true	true
true	true	false	false
true	true	true	false

P: It is a Tuesday. Q: It is raining. R: Harry will go for a run.

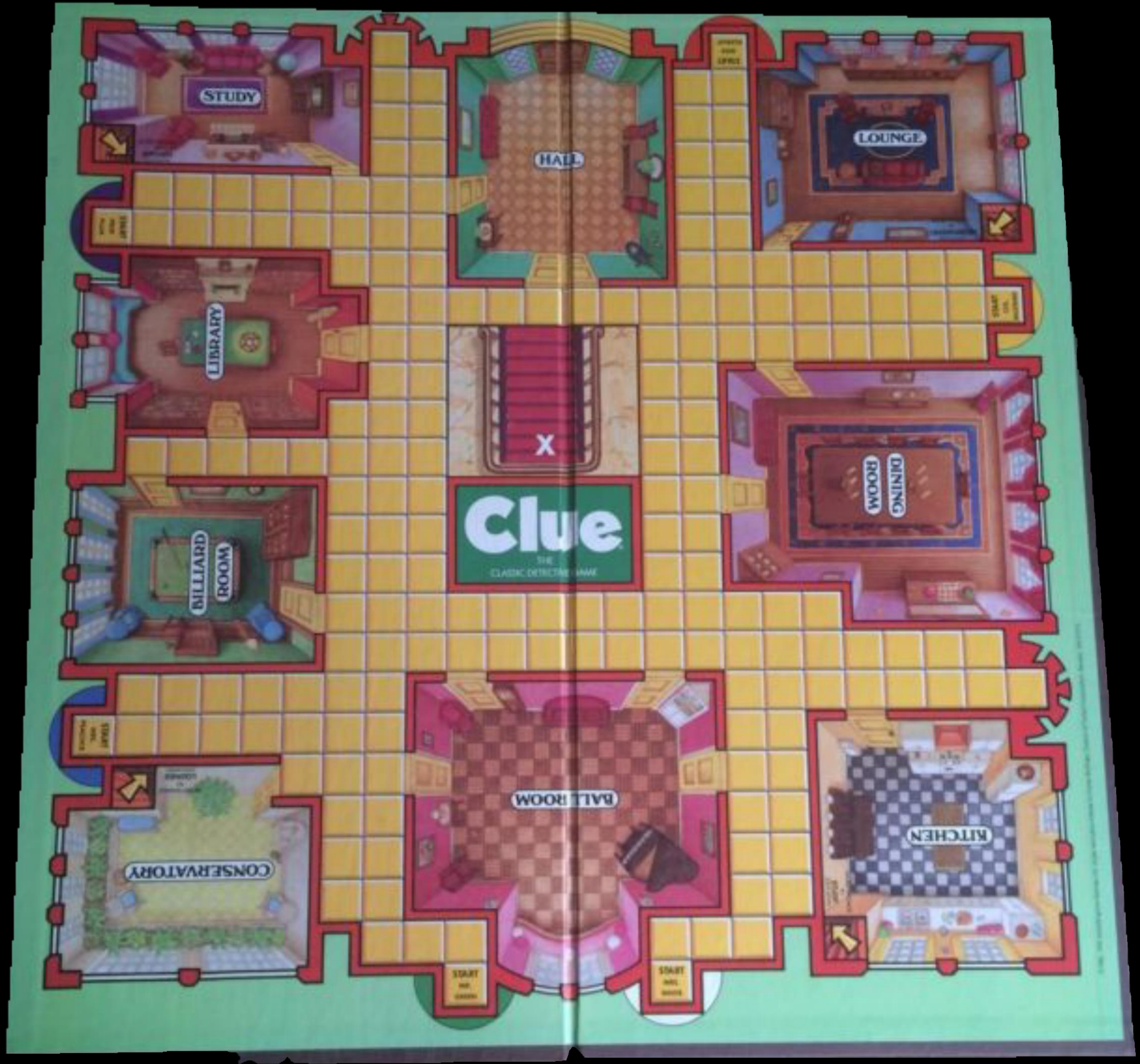
KB: $(P \land \neg Q) \rightarrow R$ P $\neg Q$

Query: R

P	Q	R	KB
false	false	false	false
false	false	true	false
false	true	false	false
false	true	true	false
true	false	false	false
true	false	true	true
true	true	false	false
true	true	true	false

Knowledge Engineering

Clue



STUDY
HALL
LOUNGE
LIBRARY
DINING ROOM
BILLIARD ROOM
Clue
THE CLASSIC DETECTIVE GAME
CONSERVATORY
BALLROOM
KITCHEN

Clue

People	Rooms	Weapons
Col. Mustard	Ballroom	Knife
Prof. Plum	Kitchen	Revolver
Ms. Scarlet	Library	Wrench

Clue

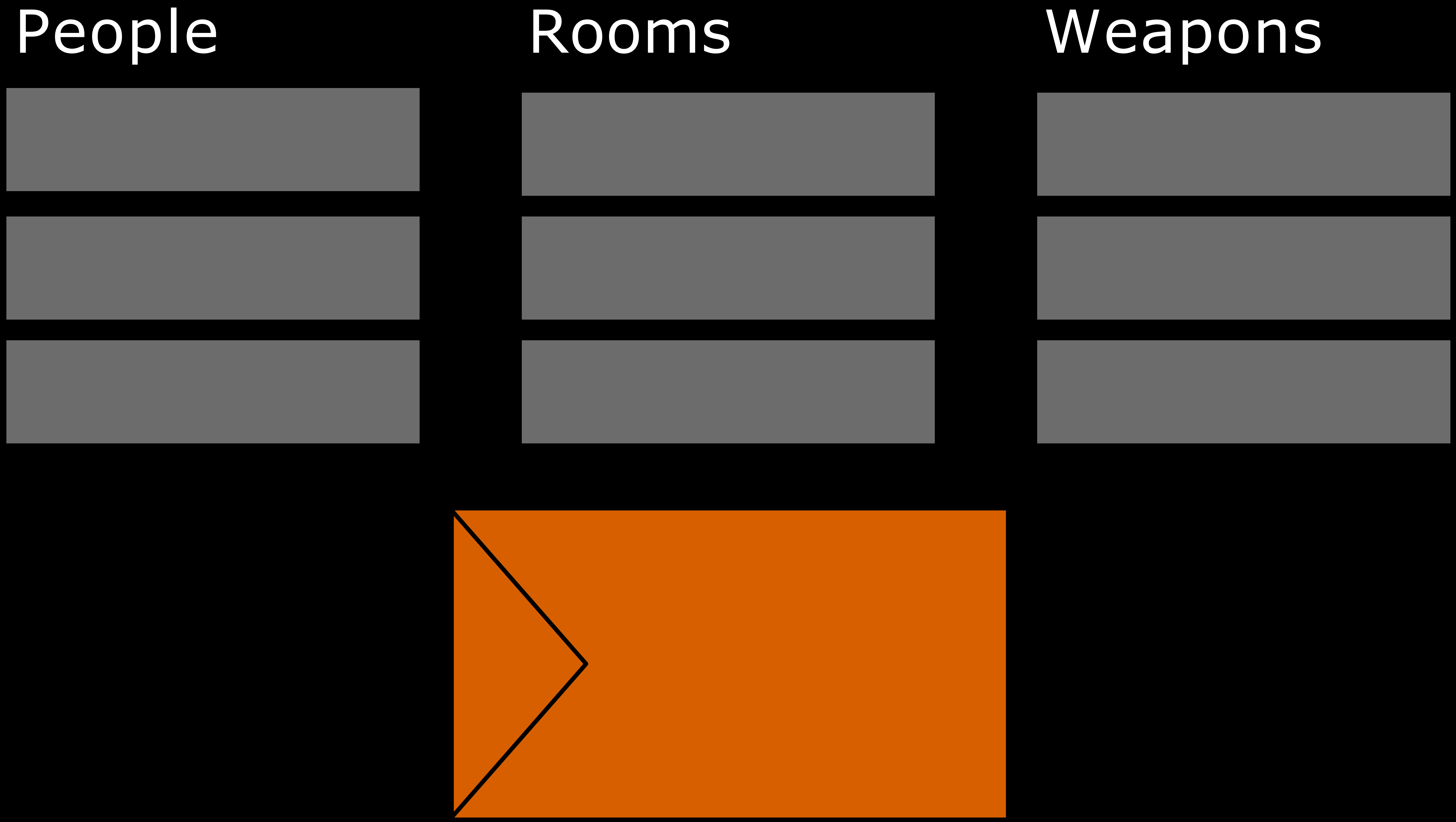
People	Rooms	Weapons

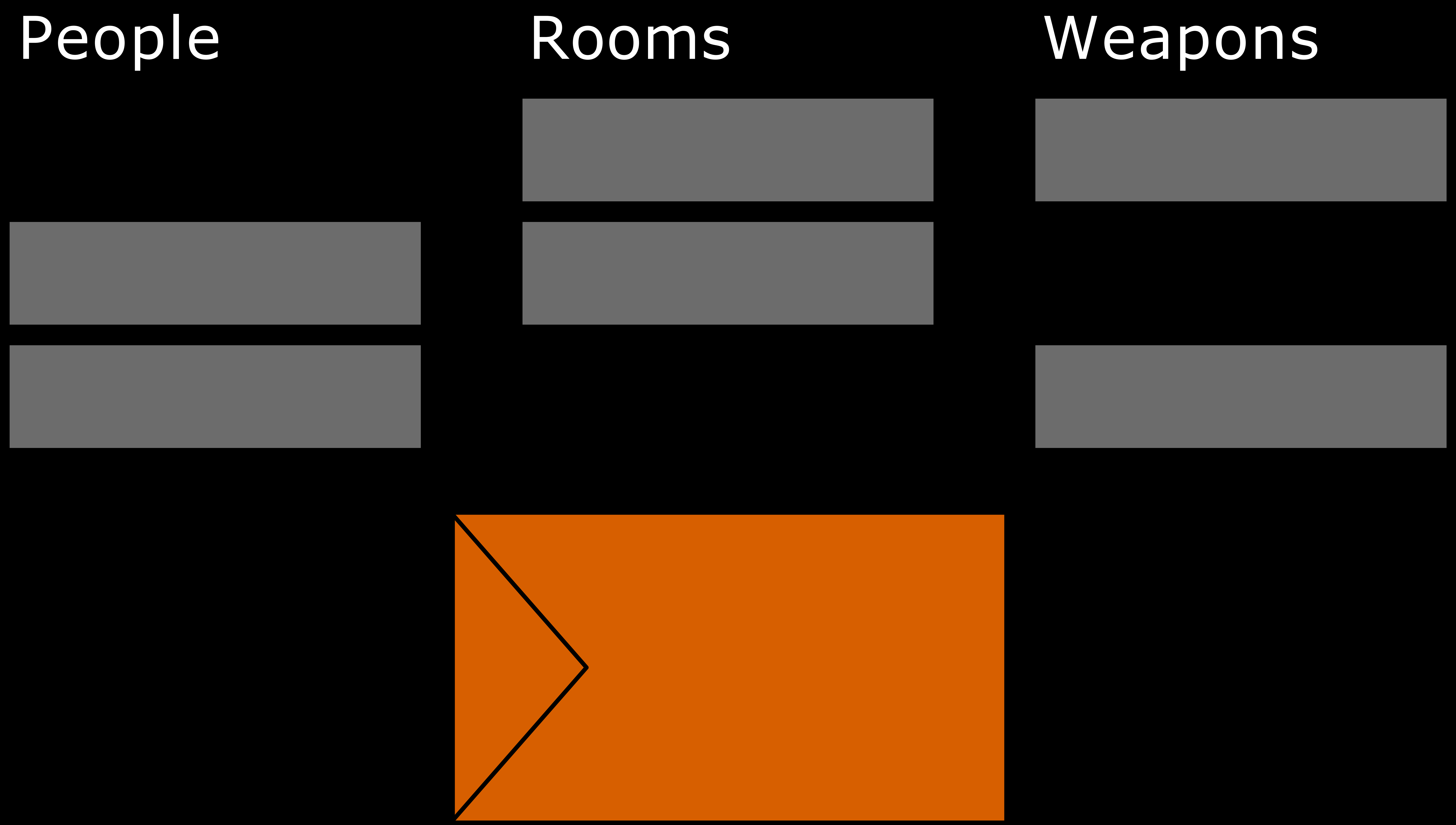
Clue

People

Rooms

Weapons


People
Rooms
Weapons


People
Rooms
Weapons

Clue

Propositional Symbols

mustard	*ballroom*	*knife*
plum	*kitchen*	*revolver*
scarlet	*library*	*wrench*

Clue

$$(mustard \lor plum \lor scarlet)$$

$$(ballroom \lor kitchen \lor library)$$

$$(knife \lor revolver \lor wrench)$$

$$\neg plum$$

$$\neg mustard \lor \neg library \lor \neg revolver$$

Logic Puzzles

- Gilderoy, Minerva, Pomona and Horace each belong to a different one of the four houses: Gryffindor, Hufflepuff, Ravenclaw, and Slytherin House.
- Gilderoy belongs to Gryffindor or Ravenclaw.
- Pomona does not belong in Slytherin.
- Minerva belongs to Gryffindor.

Logic Puzzles

Propositional Symbols

GilderoyGryffindor
GilderoyHufflepuff
GilderoyRavenclaw
GilderoySlytherin

MinervaGryffindor
MinervaHufflepuff
MinervaRavenclaw
MinervaSlytherin

PomonaGryffindor
PomonaHufflepuff
PomonaRavenclaw
PomonaSlytherin

HoraceGryffindor
HoraceHufflepuff
HoraceRavenclaw
HoraceSlytherin

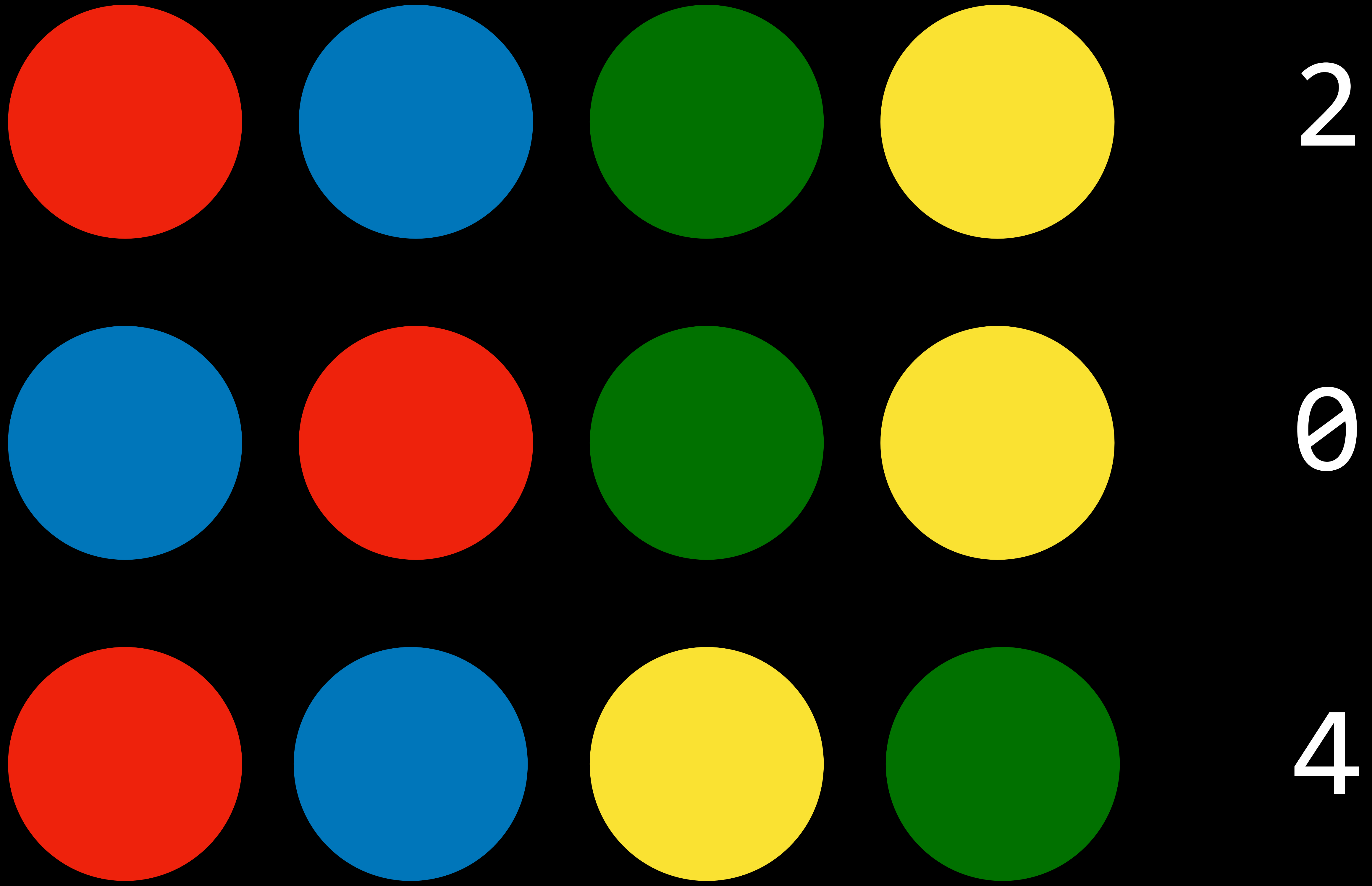
Logic Puzzles

$$(PomonaSlytherin \rightarrow \neg PomonaHufflepuff)$$

$$(MinervaRavenclaw \rightarrow \neg GilderoyRavenclaw)$$

$$(GilderoyGryffindor \lor GilderoyRavenclaw)$$

Mastermind



Inference Rules

Modus Ponens

If it is raining, then Harry is inside.

It is raining.

Harry is inside.

Modus Ponens

$$\frac{\alpha \to \beta \qquad \alpha}{\beta}$$

And Elimination

Harry is friends with Ron and Hermione.

Harry is friends with Hermione.

And Elimination

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

Double Negation Elimination

It is not true that Harry did not pass the test.

Harry passed the test.

Double Negation Elimination

$$\frac{\neg(\neg\alpha)}{\alpha}$$

Implication Elimination

If it is raining, then Harry is inside.

It is not raining or Harry is inside.

Implication Elimination

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

Biconditional Elimination

It is raining if and only if Harry is inside.

If it is raining, then Harry is inside,
and if Harry is inside, then it is raining.

Biconditional Elimination

$$\frac{\alpha \leftrightarrow \beta}{(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)}$$

De Morgan's Law

It is not true that both
Harry and Ron passed the test.

Harry did not pass the test
or Ron did not pass the test.

De Morgan's Law

$$\frac{\neg(\alpha \land \beta)}{\neg\alpha \lor \neg\beta}$$

De Morgan's Law

It is not true that
Harry or Ron passed the test.

Harry did not pass the test
and Ron did not pass the test.

De Morgan's Law

$$\frac{\neg(\alpha \vee \beta)}{\neg\alpha \wedge \neg\beta}$$

Distributive Property

$$\frac{(\alpha \wedge (\beta \vee \gamma))}{(\alpha \wedge \beta) \vee (\alpha \wedge \gamma)}$$

Distributive Property

$$\frac{(\alpha \vee (\beta \wedge \gamma))}{(\alpha \vee \beta) \wedge (\alpha \vee \gamma)}$$

Search Problems

- initial state
- actions
- transition model
- goal test
- path cost function

Theorem Proving

- initial state: starting knowledge base
- actions: inference rules
- transition model: new knowledge base after inference
- goal test: check statement we're trying to prove
- path cost function: number of steps in proof

Resolution

(Ron is in the Great Hall) $\vee$ (Hermione is in the library)

Ron is not in the Great Hall

Hermione is in the library

$$P \vee Q$$

$$\neg P$$

$$Q$$

$$P \vee Q_1 \vee Q_2 \vee \ldots \vee Q_n$$

$$\neg P$$

$$Q_1 \vee Q_2 \vee \ldots \vee Q_n$$

(Ron is in the Great Hall) $\vee$ (Hermione is in the library)

(Ron is not in the Great Hall) $\vee$ (Harry is sleeping)

(Hermione is in the library) $\vee$ (Harry is sleeping)

$$\frac{\begin{array}{c} P \lor Q \\ \neg P \lor R \end{array}}{Q \lor R}$$

$$P \vee Q_1 \vee Q_2 \vee \ldots \vee Q_n$$

$$\neg P \vee R_1 \vee R_2 \vee \ldots \vee R_m$$

$$\overline{}$$

$$Q_1 \vee Q_2 \vee \ldots \vee Q_n \vee R_1 \vee R_2 \vee \ldots \vee R_m$$

clause

a disjunction of literals

e.g. $P \vee Q \vee R$

conjunctive normal form

logical sentence that is a conjunction of clauses

e.g. $(A \vee B \vee C) \wedge (D \vee \neg E) \wedge (F \vee G)$

Conversion to CNF

- Eliminate biconditionals
  - turn $(\alpha \leftrightarrow \beta)$ into $(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$
- Eliminate implications
  - turn $(\alpha \rightarrow \beta)$ into $\neg\alpha \vee \beta$
- Move $\neg$ inwards using De Morgan's Laws
  - e.g. turn $\neg(\alpha \wedge \beta)$ into $\neg\alpha \vee \neg\beta$
- Use distributive law to distribute $\vee$ wherever possible

Conversion to CNF

$$(P \lor Q) \to R$$

$$\neg(P \lor Q) \lor R$$ eliminate implication

$$(\neg P \land \neg Q) \lor R$$ De Morgan's Law

$$(\neg P \lor R) \land (\neg Q \lor R)$$ distributive law

Inference by Resolution

$$
\frac{\begin{array}{c} P \lor Q \\ \neg P \lor R \end{array}}{(Q \lor R)}
$$

$$P \vee Q \vee S$$

$$\neg P \vee R \vee S$$

$$(Q \vee S \vee R \vee S)$$

$$P \vee Q \vee S$$

$$\neg P \vee R \vee S$$

$$(Q \vee R \vee S)$$

$$\frac{P \qquad \neg P}{(\,)}$$

Inference by Resolution

- To determine if $\mathrm{KB} \vDash \alpha$:
    - Check if $(\mathrm{KB} \wedge \neg\alpha)$ is a contradiction?
        - If so, then $\mathrm{KB} \vDash \alpha$.
        - Otherwise, no entailment.

Inference by Resolution

- To determine if $\mathrm{KB} \vDash \alpha$:
  - Convert $(\mathrm{KB} \wedge \neg\alpha)$ to Conjunctive Normal Form.
  - Keep checking to see if we can use resolution to produce a new clause.
    - If ever we produce the **empty** clause (equivalent to False), we have a contradiction, and $\mathrm{KB} \vDash \alpha$.
    - Otherwise, if we can't add new clauses, no entailment.

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

$(A \vee B)$ $\quad (\neg B \vee C)$ $\quad (\neg C)$ $\quad (\neg A)$

Inference by Resolution

Does $(A \lor B) \land (\neg B \lor C) \land (\neg C)$ entail A?

$$(A \lor B) \land (\neg B \lor C) \land (\neg C) \land (\neg A)$$

$(A \lor B) \quad (\neg B \lor C) \quad (\neg C) \quad (\neg A)$

Inference by Resolution

Does $(A \lor B) \land (\neg B \lor C) \land (\neg C)$ entail A?

$$(A \lor B) \land (\neg B \lor C) \land (\neg C) \land (\neg A)$$

$(A \lor B)$ $\quad$ $\underline{(\neg B \lor C)}$ $\quad$ $\underline{(\neg C)}$ $\quad$ $(\neg A)$ $\quad$ $(\neg B)$

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

$(A \vee B)$ $(\neg B \vee C)$ $(\neg C)$ $(\neg A)$ $(\neg B)$

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

$\underline{(A \vee B)}$ $(\neg B \vee C)$ $(\neg C)$ $(\neg A)$ $\underline{(\neg B)}$

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

$(A \vee B)$ $(\neg B \vee C)$ $(\neg C)$ $(\neg A)$ $(\neg B)$ (A)

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

$(A \vee B)$ $(\neg B \vee C)$ $(\neg C)$ $(\neg A)$ $(\neg B)$ (A)

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

$(A \vee B)$ $\quad (\neg B \vee C)$ $\quad (\neg C)$ $\quad \underline{(\neg A)}$ $\quad (\neg B)$ $\quad \underline{(A)}$

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

$(A \vee B)$ $(\neg B \vee C)$ $(\neg C)$ $\underline{(\neg A)}$ $(\neg B)$ $\underline{(A)}$ $(\,)$

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

$(A \vee B)$ $(\neg B \vee C)$ $(\neg C)$ $(\neg A)$ $(\neg B)$ (A) $(\;)$

First-Order Logic

Propositional Logic

Propositional Symbols

MinervaGryffindor

MinervaHufflepuff

MinervaRavenclaw

MinervaSlytherin

...

First-Order Logic

Constant Symbol

Minerva
Pomona
Horace
Gilderoy
Gryffindor
Hufflepuff
Ravenclaw
Slytherin

Predicate Symbol

Person
House
BelongsTo

First-Order Logic

$Person(Minerva)$ Minerva is a person.

$House(Gryffindor)$ Gryffindor is a house.

$\neg House(Minerva)$ Minerva is not a house.

$BelongsTo(Minerva, Gryffindor)$

Minerva belongs to Gryffindor.

Universal Quantification

Universal Quantification

$$\forall x.\ BelongsTo(x, Gryffindor) \rightarrow \neg BelongsTo(x, Hufflepuff)$$

For all objects x, if x belongs to Gryffindor,
then x does not belong to Hufflepuff.

Anyone in Gryffindor is not in Hufflepuff.

Existential Quantification

Existential Quantification

$$\exists x.\ House(x) \land BelongsTo(Minerva, x)$$

There exists an object x such that
x is a house and Minerva belongs to x.

Minerva belongs to a house.

Existential Quantification

$$\forall x.\ Person(x) \rightarrow (\exists y.\ House(y) \land BelongsTo(x, y))$$

For all objects x, if x is a person, then
there exists an object y such that
y is a house and x belongs to y.

Every person belongs to a house.