

7. Logical Agents

7.1 Suppose the agent has progressed to the point shown in Figure [wumpus-seq35-figure\(a\)](#), page [wumpus-seq35-figure](#), having perceived nothing in [1,1], a breeze in [2,1], and a stench in [1,2], and is now concerned with the contents of [1,3], [2,2], and [3,1]. Each of these can contain a pit, and at most one can contain a wumpus. Following the example of Figure [wumpus-entailment-figure](#), construct the set of possible worlds. (You should find 32 of them.) Mark the worlds in which the KB is true and those in which each of the following sentences is true:

α_2 = "There is no pit in [2,2]."

α_3 = "There is a wumpus in [1,3]."

Hence show that $\{KB\} \models \alpha_2$ and $\{KB\} \models \alpha_3$.

7.2 (Adapted from @Barwise+Etchemendy:1993 .) Given the following, can you prove that the unicorn is mythical? How about magical? Horned?

If the unicorn is mythical, then it is immortal, but if it is not mythical, then it is a mortal mammal. If the unicorn is either immortal or a mammal, then it is horned. The unicorn is magical if it is horned.

7.3 [truth-value-exercise] Consider the problem of deciding whether a propositional logic sentence is true in a given model.

1. Write a recursive algorithm PL-True?(s, m) that returns true if and only if the sentence s is true in the model m (where m assigns a truth value for every symbol in s). The algorithm should run in time linear in the size of the sentence. (Alternatively, use a version of this function from the online code repository.)
2. Give three examples of sentences that can be determined to be true or false in a *partial* model that does not specify a truth value for some of the symbols.
3. Show that the truth value (if any) of a sentence in a partial model cannot be determined efficiently in general.
4. Modify your algorithm so that it can sometimes judge truth from partial models, while retaining its recursive structure and linear run time. Give three examples of sentences whose truth in a partial model is *not* detected by your algorithm.
5. Investigate whether the modified algorithm makes TT-Entails? more efficient.

7.4 Which of the following are correct?

1. $\{False\} \models \{True\}$.
2. $\{True\} \models \{False\}$.
3. $(A \wedge B) \models (A \rightarrow B)$.
4. $A \rightarrow B \models A \vee B$.
5. $A \rightarrow B \models \neg A \vee B$.
6. $(A \wedge B) \rightarrow C \models (A \rightarrow C) \vee (B \rightarrow C)$.
7. $C \vee (\neg A \wedge \neg B) \equiv ((A \rightarrow C) \wedge (B \rightarrow C))$.
8. $(A \vee B) \wedge (\neg C \vee \neg D \vee E) \models (A \vee B)$.
9. $(A \vee B) \wedge (\neg C \vee \neg D \vee E) \models (A \vee B) \wedge (\neg D \vee E)$.
10. $(A \vee B) \wedge (A \rightarrow B)$ is satisfiable.
11. $(A \rightarrow B) \wedge (\neg A \vee B)$ is satisfiable.
12. $(A \rightarrow B) \rightarrow C$ has the same number of models as $(A \rightarrow B)$ for any fixed set of proposition symbols that includes A, B, C.

7.5 Which of the following are correct?

1. $\{False\} \models \{True\}$.
2. $\{True\} \models \{False\}$.
3. $(A \wedge B) \models (A \rightarrow B)$.
4. $A \rightarrow B \models A \vee B$.
5. $A \rightarrow B \models \neg A \vee B$.
6. $(A \vee B) \wedge (\neg C \vee \neg D \vee E) \models (A \vee B \vee C) \wedge (B \wedge C \wedge D \rightarrow E)$.
7. $(A \vee B) \wedge (\neg C \vee \neg D \vee E) \models (A \vee B) \wedge (\neg D \vee E)$.
8. $(A \vee B) \wedge (A \rightarrow B)$ is satisfiable.
9. $(A \wedge B) \rightarrow C \models (A \rightarrow C) \vee (B \rightarrow C)$.
10. $C \vee (\neg A \wedge \neg B) \equiv ((A \rightarrow C) \wedge (B \rightarrow C))$.
11. $(A \rightarrow B) \wedge (\neg A \vee B)$ is satisfiable.
12. $(A \rightarrow B) \rightarrow C$ has the same number of models as $(A \rightarrow B)$ for any fixed set of proposition symbols that includes A, B, C.

7.6 [deduction-theorem-exercise] Prove each of the following assertions:

1. α is valid if and only if $\{True\} \models \alpha$.
2. For any α , $\{False\} \models \alpha$.
3. $\alpha \rightarrow \beta$ is valid if and only if the sentence $\alpha \rightarrow \beta$ is valid.
4. $\alpha \equiv \beta$ is valid if and only if the sentence $\alpha \leftrightarrow \beta$ is valid.

5. $\alpha \models \beta$ if and only if the sentence $(\alpha \wedge \neg \beta)$ is unsatisfiable.

7.7 Prove, or find a counterexample to, each of the following assertions:

1. If $\alpha \models \gamma$ or $\beta \models \gamma$ (or both) then $(\alpha \wedge \beta) \models \gamma$.
2. If $(\alpha \wedge \beta) \models \gamma$ then $\alpha \models \gamma$ or $\beta \models \gamma$ (or both).
3. If $\alpha \models (\beta \vee \gamma)$ then $\alpha \models \beta$ or $\alpha \models \gamma$ (or both).

7.8 Prove, or find a counterexample to, each of the following assertions:

1. If $\alpha \models \gamma$ or $\beta \models \gamma$ (or both) then $(\alpha \wedge \beta) \models \gamma$.
2. If $\alpha \models (\beta \wedge \gamma)$ then $\alpha \models \beta$ and $\alpha \models \gamma$.
3. If $\alpha \models (\beta \vee \gamma)$ then $\alpha \models \beta$ or $\alpha \models \gamma$ (or both).

7.9 Consider a vocabulary with only four propositions, A , B , C , and D . How many models are there for the following sentences?

1. $B \vee C$.
2. $\neg A \vee \neg B \vee \neg C \vee \neg D$.
3. $(A \rightarrow B) \wedge A \wedge \neg B \wedge C \wedge D$.

7.10 We have defined four binary logical connectives.

1. Are there any others that might be useful?
2. How many binary connectives can there be?
3. Why are some of them not very useful?

7.11 [logical-equivalence-exercise] Using a method of your choice, verify each of the equivalences in Table [logical-equivalence-table] (page [logical-equivalence-table](#)).

7.12 [propositional-validity-exercise] Decide whether each of the following sentences is valid, unsatisfiable, or neither. Verify your decisions using truth tables or the equivalence rules of Table [logical-equivalence-table] (page [logical-equivalence-table](#)).

1. $\{Smoke\} \vdash \{Smoke\}$
2. $\{Smoke\} \vdash \{Fire\}$
3. $(\{Smoke\} \vdash \{Fire\}) \vdash (\neg \{Smoke\} \vdash \neg \{Fire\})$
4. $\{Smoke\} \vee \{Fire\} \vee \neg \{Fire\}$
5. $((\{Smoke\} \wedge \{Heat\}) \vdash \{Fire\})$

$\vdash ((\{Smoke\} \vdash \{Fire\}) \vee (\{Heat\} \vdash \{Fire\}))$

6. $(\{Smoke\} \vdash \{Fire\}) \vdash ((\{Smoke\} \wedge \{Heat\}) \vdash \{Fire\})$
7. $\{Big\} \vee \{Dumb\} \vee (\{Big\} \vdash \{Dumb\})$

7.13 [propositional-validity-exercise] Decide whether each of the following sentences is valid, unsatisfiable, or neither. Verify your decisions using truth tables or the equivalence rules of Table [logical-equivalence-table] (page [logical-equivalence-table](#)).

1. $\{Smoke\} \vdash \{Smoke\}$
2. $\{Smoke\} \vdash \{Fire\}$
3. $(\{Smoke\} \vdash \{Fire\}) \vdash (\neg \{Smoke\} \vdash \neg \{Fire\})$
4. $\{Smoke\} \vee \{Fire\} \vee \neg \{Fire\}$
5. $((\{Smoke\} \wedge \{Heat\}) \vdash \{Fire\})$

$\vdash ((\{Smoke\} \vdash \{Fire\}) \vee (\{Heat\} \vdash \{Fire\}))$

6. $\{Big\} \vee \{Dumb\} \vee (\{Big\} \vdash \{Dumb\})$
7. $(\{Big\} \wedge \{Dumb\}) \vee \neg \{Dumb\}$

7.14 [cnf-proof-exercise] Any propositional logic sentence is logically equivalent to the assertion that each possible world in which it would be false is not the case. From this observation, prove that any sentence can be written in CNF.

7.15 Use resolution to prove the sentence $\neg A \wedge \neg B$ from the clauses in Exercise [convert-clausal-exercise](#).

7.16 [inf-exercise] This exercise looks into the relationship between clauses and implication sentences.

1. Show that the clause $(\neg P_1 \vee \dots \vee \neg P_m \vee Q)$ is logically equivalent to the implication sentence $(P_1 \wedge \dots \wedge P_m) \rightarrow Q$.
2. Show that every clause (regardless of the number of positive literals) can be written in the form $(P_1 \wedge \dots \wedge P_m) \rightarrow (Q_1 \vee \dots \vee Q_n)$, where the P 's and Q 's are proposition symbols. A knowledge base consisting of such sentences is in implicative normal form or **Kowalski form** @Kowalski:1979.
3. Write down the full resolution rule for sentences in implicative normal form.

7.17 According to some political pundits, a person who is radical (R) is electable (E) if he/she is conservative (C), but otherwise is not electable.

1. Which of the following are correct representations of this assertion?

- A. $(R \wedge E) \iff C$
 B. $R \rightarrow (E \iff C)$
 C. $R \rightarrow ((C \rightarrow E) \vee \neg E)$
2. Which of the sentences in (a) can be expressed in Horn form?

7.18 This question considers representing satisfiability (SAT) problems as CSPs.

- Draw the constraint graph corresponding to the SAT problem $(\neg X_1 \vee X_2) \wedge (\neg X_2 \vee X_3) \wedge \dots \wedge (\neg X_{n-1} \vee X_n)$ for the particular case $n = 5$.
- How many solutions are there for this general SAT problem as a function of n ?
- Suppose we apply {Backtracking-Search} (page [backtracking-search-algorithm](#)) to find *all* solutions to a SAT CSP of the type given in (a). (To find *all* solutions to a CSP, we simply modify the basic algorithm so it continues searching after each solution is found.) Assume that variables are ordered X_1, \dots, X_n and $\{false\}$ is ordered before $\{true\}$. How much time will the algorithm take to terminate? (Write an $O(\cdot)$ expression as a function of n .)
- We know that SAT problems in Horn form can be solved in linear time by forward chaining (unit propagation). We also know that every tree-structured binary CSP with discrete, finite domains can be solved in time linear in the number of variables (Section [csp-structure-section](#)). Are these two facts connected? Discuss.

7.19 This question considers representing satisfiability (SAT) problems as CSPs.

- Draw the constraint graph corresponding to the SAT problem $(\neg X_1 \vee X_2) \wedge (\neg X_2 \vee X_3) \wedge \dots \wedge (\neg X_{n-1} \vee X_n)$ for the particular case $n = 4$.
- How many solutions are there for this general SAT problem as a function of n ?
- Suppose we apply {Backtracking-Search} (page [backtracking-search-algorithm](#)) to find *all* solutions to a SAT CSP of the type given in (a). (To find *all* solutions to a CSP, we simply modify the basic algorithm so it continues searching after each solution is found.) Assume that variables are ordered X_1, \dots, X_n and $\{false\}$ is ordered before $\{true\}$. How much time will the algorithm take to terminate? (Write an $O(\cdot)$ expression as a function of n .)
- We know that SAT problems in Horn form can be solved in linear time by forward chaining (unit propagation). We also know that every tree-structured binary CSP with discrete, finite domains can be solved in time linear in the number of variables (Section [csp-structure-section](#)). Are these two facts connected? Discuss.

7.20 Explain why every nonempty propositional clause, by itself, is satisfiable. Prove rigorously that every set of five 3-SAT clauses is satisfiable, provided that each clause mentions exactly three distinct variables. What is the smallest set of such clauses that is unsatisfiable? Construct such a set.

7.21 A propositional 2-CNF expression is a conjunction of clauses, each containing *exactly* 2 literals, e.g., $(A \vee B) \wedge (\neg A \vee C) \wedge (\neg B \vee D) \wedge (\neg C \vee G) \wedge (\neg D \vee G)$.

- Prove using resolution that the above sentence entails G .
- Two clauses are *semantically distinct* if they are not logically equivalent. How many semantically distinct 2-CNF clauses can be constructed from n proposition symbols?
- Using your answer to (b), prove that propositional resolution always terminates in time polynomial in n given a 2-CNF sentence containing no more than n distinct symbols.
- Explain why your argument in (c) does not apply to 3-CNF.

7.22 Prove each of the following assertions:

- Every pair of propositional clauses either has no resolvents, or all their resolvents are logically equivalent.
- There is no clause that, when resolved with itself, yields (after factoring) the clause $(\neg P \vee \neg Q)$.
- If a propositional clause C can be resolved with a copy of itself, it must be logically equivalent to True .

7.23 Consider the following sentence: $((\text{Food} \rightarrow \text{Party}) \vee (\text{Drinks} \rightarrow \text{Party})) \rightarrow (\text{Food} \wedge \text{Drinks} \rightarrow \text{Party})$.

- Determine, using enumeration, whether this sentence is valid, satisfiable (but not valid), or unsatisfiable.
- Convert the left-hand and right-hand sides of the main implication into CNF, showing each step, and explain how the results confirm your answer to (a).
- Prove your answer to (a) using resolution.

7.24 [dnf-exercise] A sentence is in disjunctive normal form (DNF) if it is the disjunction of conjunctions of literals. For example, the sentence $(A \wedge B \wedge \neg C) \vee (\neg A \wedge C) \vee (B \wedge \neg C)$ is in DNF.

- Any propositional logic sentence is logically equivalent to the assertion that some possible world in which it would be true is in fact the case. From this observation, prove that any sentence can be written in DNF.
- Construct an algorithm that converts any sentence in propositional logic into DNF. (*Hint*: The algorithm is similar to the algorithm for conversion to CNF even in Section [pl-resolution-section](#).)
- Construct a simple algorithm that takes as input a sentence in DNF and returns a satisfying assignment if one exists, or reports that no satisfying assignment exists.
- Apply the algorithms in (b) and (c) to the following set of sentences:

$A \rightarrow B$
 $B \rightarrow C$
 $C \rightarrow A$

1. Since the algorithm in (b) is very similar to the algorithm for conversion to CNF, and since the algorithm in (c) is much simpler than any algorithm for solving a set of sentences in CNF, why is this technique not used in automated reasoning?

7.25 [convert-clausal-exercise] Convert the following set of sentences to clausal form.

S1: $A \rightarrow (B \vee E)$.
S2: $E \rightarrow D$.
S3: $C \wedge F \rightarrow \neg B$.
S4: $E \rightarrow B$.
S5: $B \rightarrow F$.
S6: $B \rightarrow C$

Give a trace of the execution of DPLL on the conjunction of these clauses.

7.26 [convert-clausal-exercise] Convert the following set of sentences to clausal form.

S1: $A \rightarrow (C \vee E)$.
S2: $E \rightarrow D$.
S3: $B \wedge F \rightarrow \neg C$.
S4: $E \rightarrow C$.
S5: $C \rightarrow F$.
S6: $C \rightarrow B$

Give a trace of the execution of DPLL on the conjunction of these clauses.

7.27 Is a randomly generated 4-CNF sentence with n symbols and m clauses more or less likely to be solvable than a randomly generated 3-CNF sentence with n symbols and m clauses? Explain.

7.28 [minesweeper-exercise] Minesweeper, the well-known computer game, is closely related to the wumpus world. A minesweeper world is a rectangular grid of N squares with M invisible mines scattered among them. Any square may be probed by the agent; instant death follows if a mine is probed. Minesweeper indicates the presence of mines by revealing, in each probed square, the *number* of mines that are directly or diagonally adjacent. The goal is to probe every unmined square.

- Let $X_{i,j}$ be true iff square $[i,j]$ contains a mine. Write down the assertion that exactly two mines are adjacent to $[1,1]$ as a sentence involving some logical combination of $X_{i,j}$ propositions.
- Generalize your assertion from (a) by explaining how to construct a CNF sentence asserting that k of n neighbors contain mines.
- Explain precisely how an agent can use {DPLL} to prove that a given square does (or does not) contain a mine, ignoring the global constraint that there are exactly M mines in all.
- Suppose that the global constraint is constructed from your method from part (b). How does the number of clauses depend on M and N ? Suggest a way to modify {DPLL} so that the global constraint does not need to be represented explicitly.
- Are any conclusions derived by the method in part (c) invalidated when the global constraint is taken into account?
- Give examples of configurations of probe values that induce *long-range dependencies* such that the contents of a given unprobed square would give information about the contents of a far-distant square. (*Hint*: consider an $N \times 1$ board.)

7.29 [known-literal-exercise] How long does it take to prove $\{KB\}^\alpha$ using {DPLL} when α is a literal *already contained in* $\{KB\}$? Explain.

7.30 [dpll-fc-exercise] Trace the behavior of {DPLL} on the knowledge base in Figure [pl-horn-example-figure](#) when trying to prove Q , and compare this behavior with that of the forward-chaining algorithm.

7.31 Write a successor-state axiom for the $\{Locked\}$ predicate, which applies to doors, assuming the only actions available are $\{Lock\}$ and $\{Unlock\}$.

7.32 Discuss what is meant by *optimal* behavior in the wumpus world. Show that the {Hybrid-Wumpus-Agent} is not optimal, and suggest ways to improve it.

7.33 Suppose an agent inhabits a world with two states, S and $\neg S$, and can do exactly one of two actions, a and b . Action a does nothing and action b flips from one state to the other. Let S^t be the proposition that the agent is in state S at time t , and let a^t be the proposition that the agent does action a at time t (similarly for b^t).

- Write a successor-state axiom for S^{t+1} .
- Convert the sentence in (a) into CNF.
- Show a resolution refutation proof that if the agent is in $\neg S$ at time t and does a , it will still be in $\neg S$ at time $t+1$.

7.34 [ss-axiom-exercise] Section [successor-state-section](#) provides some of the successor-state axioms required for the wumpus world. Write down axioms for all remaining fluent symbols.

7.35 [hybrid-wumpus-exercise] Modify the {Hybrid-Wumpus-Agent} to use the 1-CNF logical state estimation method described on page [1cnf-belief-state-page](#). We noted on that page that such an agent will not be able to acquire, maintain, and use more complex beliefs such as the disjunction $P_{3,1} \vee P_{2,2}$. Suggest a method for overcoming this problem by defining additional proposition symbols, and try it out in the wumpus world. Does it improve the performance of the agent?