Exercise 4: Logical Agents - Solutions Winter Semester 2022/23
Adrian Kulmburg

Problem 4.1: Model, satisfaction relation, and entailment

(*Taken from [3] question 7.4*) Which of the following statements are correct? Show its correctness by reasoning about the model satisfying each sentence.

1. $False \models True$
2. $True \models False$
3. $(A \wedge B) \models (A \Leftrightarrow B)$
4. $A \Leftrightarrow B \models A \vee B$
5. $A \Leftrightarrow B \models \neg A \vee B$

Solution:

1.
$$\begin{aligned}
& False \models True \\
& \quad \text{(Definition of entailment)} \\
\leftrightarrow\ & M(False) \subseteq M(True) \\
& \quad \text{(Model of } False \text{ is empty set)} \\
\leftrightarrow\ & \emptyset \subseteq M(True) \\
& \quad (\emptyset \text{ is subset of any set}) \\
\leftrightarrow\ & \text{Correct}
\end{aligned}$$

2.
$$\begin{aligned}
& True \models False \\
& \quad \text{(Definition of entailment)} \\
\leftrightarrow\ & M(True) \subseteq M(False) \\
& \quad \text{(Model of } False \text{ is empty set)} \\
\leftrightarrow\ & M(True) \subseteq \emptyset \\
& \quad (M(True) \text{ is the set of all possible models; hence, it is } \neq \emptyset) \\
\leftrightarrow\ & \text{Incorrect}
\end{aligned}$$

3.
$$\begin{aligned}
& (A \wedge B) \models (A \Leftrightarrow B) \\
& \quad \text{(Definition of entailment)} \\
\leftrightarrow\ & M(A \wedge B) \subseteq M(A \Leftrightarrow B) \\
& \quad \text{(A model satisfying } A \wedge B \text{ is given by a truth value assignment for which } A \wedge B \text{ is true.} \\
& \quad \text{This happens if and only if } A \text{ and } B \text{ are both true, thus } M(A \wedge B) = \{(\text{True, True})\}) \\
\leftrightarrow\ & \{(True, True)\} \subseteq M(A \Leftrightarrow B) \\
& \quad \text{(The models satisfying } A \Leftrightarrow B \text{ are } \{(True, True), (False, False)\}) \\
\leftrightarrow\ & \{(True, True)\} \subseteq \{(True, True), (False, False)\} \\
\leftrightarrow\ & \text{Correct}
\end{aligned}$$

4. $(A \Leftrightarrow B) \models (A \vee B)$

(Definition of entailment)

$\leftrightarrow$ $M(A \Leftrightarrow B) \subseteq M(A \vee B)$

(The models satisfying $A \Leftrightarrow B$ are $\{(True, True), (False, False)\}$)

$\leftrightarrow$ $\{(True, True), (False, False)\} \subseteq M(A \vee B)$

(The models satisfying $A \vee B$ are $\{(True, True), (True, False), (False, True)\}$)

$\leftrightarrow$ $\{(True, True), (False, False)\} \subseteq \{(True, True), (True, False), (False, True)\}$

$\leftrightarrow$ Incorrect

5. $(A \Leftrightarrow B) \models (\neg A \vee B)$

(Definition of entailment)

$\leftrightarrow$ $M(A \Leftrightarrow B) \subseteq M(\neg A \vee B)$

(The models satisfying $A \Leftrightarrow B$ are $\{(True, True), (False, False)\}$)

$\leftrightarrow$ $\{(True, True), (False, False)\} \subseteq M(\neg A \vee B)$

(The models satisfying $\neg A \vee B$ are $\{(True, True), (False, True), (False, False)\}$)

$\leftrightarrow$ $\{(True, True), (False, False)\} \subseteq \{(True, True), (False, True), (False, False)\}$

$\leftrightarrow$ Correct

Problem 4.2: Validity, satisfiability, and unsatisfiability

(*Exercise is adapted from [3] question 7.10.*) Recall first the definition of *validity* and *satisfiability*.

Problem 4.2.1: Prove the following two metatheorems:

1. Sentence α is valid if and only if $\alpha \equiv True$,
2. Sentence α is unsatisfiable if and only if $\alpha \equiv False$.

Solution:

1. α is valid

(Definition of validity)

$\leftrightarrow$ $M(\alpha) = \text{AllPossibleModels}$

(Since $M(True) = \text{AllPossibleModels}$)

$\leftrightarrow$ $M(\alpha) = M(True)$

(Definition of set equality)

$\leftrightarrow$ $(M(\alpha) \subseteq M(True)) \wedge (M(True) \subseteq M(\alpha))$

(Definition of entailment)

$\leftrightarrow$ $(\alpha \models True)$ and $(True \models \alpha)$

(Definition of logical equivalence)

$\leftrightarrow$ $\alpha \equiv True$

2. α is unsatisfiable

(Definition of satisfiability)

$\leftrightarrow$ $M(\alpha) = \emptyset$

(Since $M(False) = \emptyset$)

$\leftrightarrow$ $M(\alpha) = M(False)$

(Definition of set equality)

$\leftrightarrow$ $(M(\alpha) \subseteq M(False)) \wedge (M(False) \subseteq M(\alpha))$

(Definition of entailment)

$\leftrightarrow$ $(\alpha \models False)$ and $(False \models \alpha)$

(Definition of logical equivalence)

$\leftrightarrow$ $\alpha \equiv False$

Problem 4.2.2: Show whether each of the following sentences is valid, satisfiable, or unsatisfiable by using the two metatheorems above, the standard logical equivalences from the lecture, and the following three logical equivalences:

$$\alpha \vee \neg\alpha \equiv True \quad (1)$$
$$\alpha \wedge \alpha \equiv \alpha \quad (2)$$
$$\alpha \vee \alpha \equiv \alpha \quad (3)$$

1. $Smoke \Rightarrow Smoke$
2. $(Smoke \Rightarrow Fire) \Rightarrow (\neg Smoke \Rightarrow \neg Fire)$
3. $Smoke \vee Fire \vee \neg Fire$
4. $(Fire \Rightarrow Smoke) \wedge Fire \wedge \neg Smoke$

Solution:

1. $Smoke \Rightarrow Smoke$

(By implication elimination $(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta)$)

$\leftrightarrow$ $\neg Smoke \vee Smoke$

(By excluded middle rule $\alpha \vee \neg\alpha \equiv True$)

$\leftrightarrow$ $True$

Since we have shown $(Smoke \Rightarrow Smoke) \equiv True$, by Problem 4.2.1 we conclude that it is valid. Since it is valid, it must be satisfiable as well.

2. $$(Smoke \Rightarrow Fire) \Rightarrow (\neg Smoke \Rightarrow \neg Fire)$$

(By implication elimination $(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta)$)

$$\leftrightarrow \neg(Smoke \Rightarrow Fire) \vee (\neg Smoke \Rightarrow \neg Fire)$$

(By implication elimination $(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta)$)

$$\leftrightarrow \neg(\neg Smoke \vee Fire) \vee (Smoke \vee \neg Fire)$$

(By the De Morgan rule $\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta)$)

$$\leftrightarrow (Smoke \wedge \neg Fire) \vee (Smoke \vee \neg Fire)$$

(By the commutativity of $\vee$)

$$\leftrightarrow (Smoke \vee \neg Fire) \vee (Smoke \wedge \neg Fire)$$

(By distributivity of $\vee$ over $\wedge$, that is $(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma))$, with $\alpha := (Smoke \vee \neg Fire)$)

$$\leftrightarrow [Smoke \vee \neg Fire \vee Smoke] \wedge [Smoke \vee \neg Fire \vee \neg Fire]$$

(By the commutativity of $\vee$)

$$\leftrightarrow [Smoke \vee Smoke \vee \neg Fire] \wedge [Smoke \vee \neg Fire \vee \neg Fire]$$

(By the rule $(\alpha \vee \alpha) \equiv \alpha$)

$$\leftrightarrow (Smoke \vee \neg Fire) \wedge (Smoke \vee \neg Fire)$$

By the rule $(\alpha \wedge \alpha) \equiv \alpha$)

$$\leftrightarrow Smoke \vee \neg Fire$$

The models satisfying $Smoke \vee \neg Fire$ are $\{(True, False), (True, True), (False, False)\}$, in particular $(False, True)$ does not satisfy it. As a consequence, $Smoke \vee \neg Fire \not\equiv True$, showing that it is not a valid sentence. However, since there exists a model satisfying $Smoke \vee \neg Fire$, it is satisfiable.

3. $$Smoke \vee Fire \vee \neg Fire$$

(By excluded middle rule $\alpha \vee \neg\alpha \equiv True$)

$$\leftrightarrow Smoke \vee True$$

(By the rule $\alpha \vee True \equiv True$)

$$\leftrightarrow True$$

We have shown $Smoke \vee Fire \vee \neg Fire \equiv True$, therefore the sentence is valid. Since it is valid, it is also satisfiable.

4. $(Fire \Rightarrow Smoke) \wedge Fire \wedge \neg Smoke$

(By implication elimination $(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta)$)

$\leftrightarrow \quad (\neg Fire \vee Smoke) \wedge Fire \wedge \neg Smoke$

(By distributivity of $\wedge$ over $\vee$, that is $(\alpha \wedge (\beta \vee \gamma)) \equiv (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$)

$\leftrightarrow \quad ((\neg Fire \wedge Fire) \vee (Smoke \wedge Fire)) \wedge \neg Smoke$

(By the rule $(\alpha \wedge \neg\alpha) \equiv False$)

$\leftrightarrow \quad (False \vee (Smoke \wedge Fire)) \wedge \neg Smoke$

(By the rule $(\alpha \vee False) \equiv \alpha$)

$\leftrightarrow \quad Smoke \wedge Fire \wedge \neg Smoke$

(By commutativity of $\wedge$)

$\leftrightarrow \quad Smoke \wedge \neg Smoke \wedge Fire$

(By the rule $(\alpha \wedge \neg\alpha) \equiv False$)

$\leftrightarrow \quad False \wedge Fire$

(By the rule $(\alpha \wedge False) \equiv False$)

$\leftrightarrow \quad False$

This shows that the sentence is unsatisfiable. Since it is unsatisfiable, it is also not valid.

Problem 4.3: Knights and Knaves

(*Proof by inference rule.*) The following puzzle is taken from [1]. Suppose we are on an island with two types of inhabitants: "knights" who always tell the truth, and "knaves" who always lie.

> According to this problem, three of the inhabitants – A, B and C – were standing together in the garden. A stranger passed by and asked A, "Are you a knight or a knave?". A answered, but rather indistinctly, so the stranger could not make out what he said. The stranger then asked B, "What did A say?". B replied, "A said that he is a knave". At this point the third man, C, said "Don't believe B; he's lying!". The question is, what are B and C?

Model this logic puzzle by introducing three atomic propositions A, B, and C with intended interpretation that A, B, and C are knights.

Problem 4.3.1: How can you formalise the sentence A *says that* B *is a knight*?

Solution: If A is a knight (A is $True$), then A must tell the truth. Therefore, B must be a knight too (B is $True$).

If A is a knave (A is $False$), then A lies. Therefore, B must be a knave too (B is $False$).

It is true when both are true or both are false (when A and B are of the same type). The proper formulation for this sentence is therefore

$$A \Leftrightarrow B.$$

Problem 4.3.2: Assume that *Remark* represents what a person says and that we can represent it using propositional logic[1]. Additionally, assume that P could either be A, B, or C. From the previous problem, can you generalise the method to model the sentence "person P says (or replies) *Remark*"?

[1]This is a mouthful way to say that *Remark* is a propositional logic sentence.

Solution: The sentence *Remark* can be either true or false, and the same goes for the propositional variable P. Now, if P is true (P is a knight), then *Remark* has to be true as well, since knights always tell the truth. Similarly, if P is false (P is a knave), *Remark* is false. We can therefore formulate "person P says *Remark*" as

$$P \Leftrightarrow Remark.$$

Problem 4.3.3: Model the following facts which are taken from the puzzle:

1. B replies, "A said that he is a knave".
2. C says, "Don't believe B; he's lying!".

Solution:

1. $B \Leftrightarrow (A \Leftrightarrow \neg A)$
2. $C \Leftrightarrow \neg B$

Problem 4.3.4: By using the following logical equivalences

$$\begin{aligned} (X \Leftrightarrow \neg X) &\equiv False && (4) \\ (X \Leftrightarrow False) &\equiv \neg X && (5) \end{aligned}$$

and the following deduction (inference) rule

$$\frac{P \Leftrightarrow Q \qquad Q}{P}$$

deduce what B and C are.

Solution: We first consider B. Starting with $B \Leftrightarrow (A \Leftrightarrow \neg A)$, we see that

$$(B \Leftrightarrow (A \Leftrightarrow \neg A)) \equiv (B \Leftrightarrow False) \equiv \neg B.$$

This shows that $\neg B$ must be true, and therefore that B is false, so B **is a knave**.

Since we have $C \Leftrightarrow \neg B$ and $\neg B$, we can infer

$$\frac{C \Leftrightarrow \neg B \qquad \neg B}{C}$$

hence C is true, meaning that C **is a knight**.

Problem 4.4: Superman does not exist

(*Proof by resolution.*) The following text is taken from [2].

> If Superman were able and willing to prevent evil, he would do so. If Superman were unable to prevent evil, he would be impotent; if he were unwilling to prevent evil, he would be malevolent. Superman does not prevent evil. If Superman exists, he is neither impotent nor malevolent. Therefore, Superman does not exist.

Assume that we use the following variables and their meaning:

A : Superman is able to prevent evil.
W : Superman is willing to prevent evil.
I : Superman is impotent.
M : Superman is malevolent.
P : Superman prevents evil.
E : Superman exists.

Problem 4.4.1: By using the variables defined above, and assuming that we want to prove "Superman does not exist", formalise the rest of the facts from the text.

Solution:

1. If Superman were able and willing to prevent evil, he would do so:
$$A \wedge W \Rightarrow P.$$
2. If Superman were unable to prevent evil, he would be impotent:
$$\neg A \Rightarrow I.$$
3. If he were unwilling to prevent evil, he would be malevolent:
$$\neg W \Rightarrow M.$$
4. Superman does not prevent evil:
$$\neg P$$
5. If Superman exists, he is neither impotent nor malevolent:
$$E \Rightarrow (\neg I \wedge \neg M).$$

Problem 4.4.2: Identify which sentences belong to the knowledge base KB, and which sentence we want to deduce. Recall the resolution approach for propositional logic. What shall we do next?

Solution: The knowledge base KB is the conjunction of all the sentences above. What we want to prove using the knowledge base is $\neg E$ (Superman does not exist). The resolution principle is based on the following theorem:

Theorem 1. *For any two propositional sentences α and β, $\alpha \models \beta$ if and only if $\alpha \wedge \neg\beta$ is unsatisfiable.*

Notice that this is exactly proof by contradiction (*reductio ad absurdum*)!!

What we should do next is to find the clausal representation (CNF) of $KB \wedge E$, so that we may apply the resolution principle. If we prove that $KB \wedge E$ is unsatisfiable, then by Theorem 1 we would have proven that $KB \models \neg E$, demonstrating that Superman does not exist under the given assumptions.

The CNF can be computed as follows:

$$\begin{aligned}
& [(A \wedge W) \Rightarrow P] \wedge [(\neg A \Rightarrow I)] \wedge [\neg W \Rightarrow M] \wedge [\neg P] \wedge [E \Rightarrow (\neg I \wedge \neg M)] \wedge [E] \\
& \quad \text{(Replace } \alpha \Rightarrow \beta \text{ by } \neg\alpha \vee \beta\text{)} \\
\leftrightarrow \; & [\neg(A \wedge W) \vee P] \wedge [A \vee I] \wedge [W \vee M] \wedge [\neg P] \wedge [\neg E \vee (\neg I \wedge \neg M)] \wedge [E] \\
& \quad \text{(Replace } \neg(\alpha \wedge \beta) \text{ by } \neg\alpha \vee \neg\beta\text{)} \\
\leftrightarrow \; & [\neg A \vee \neg W \vee P] \wedge [A \vee I] \wedge [W \vee M] \wedge [\neg P] \wedge [\neg E \vee (\neg I \wedge \neg M)] \wedge [E] \\
& \quad \text{(Replace } \alpha \vee (\beta \wedge \gamma) \text{ by } (\alpha \vee \beta) \wedge (\alpha \vee \gamma)\text{)} \\
\leftrightarrow \; & [\neg A \vee \neg W \vee P] \wedge [A \vee I] \wedge [W \vee M] \wedge [\neg P] \wedge [\neg E \vee \neg I] \wedge [\neg E \vee \neg M] \wedge [E]
\end{aligned}$$

The set of clauses is therefore:

1. $\neg A \vee \neg W \vee P$

2. $A \vee I$
3. $W \vee M$
4. $\neg P$
5. $\neg E \vee \neg I$
6. $\neg E \vee \neg M$
7. E

Problem 4.4.3: Prove diagramatically with the resolution approach that "Superman does not exist".

Solution: See Figure 1. At the end of the algorithm, we end up with an empty clause, showing that $KB \models \neg E$.

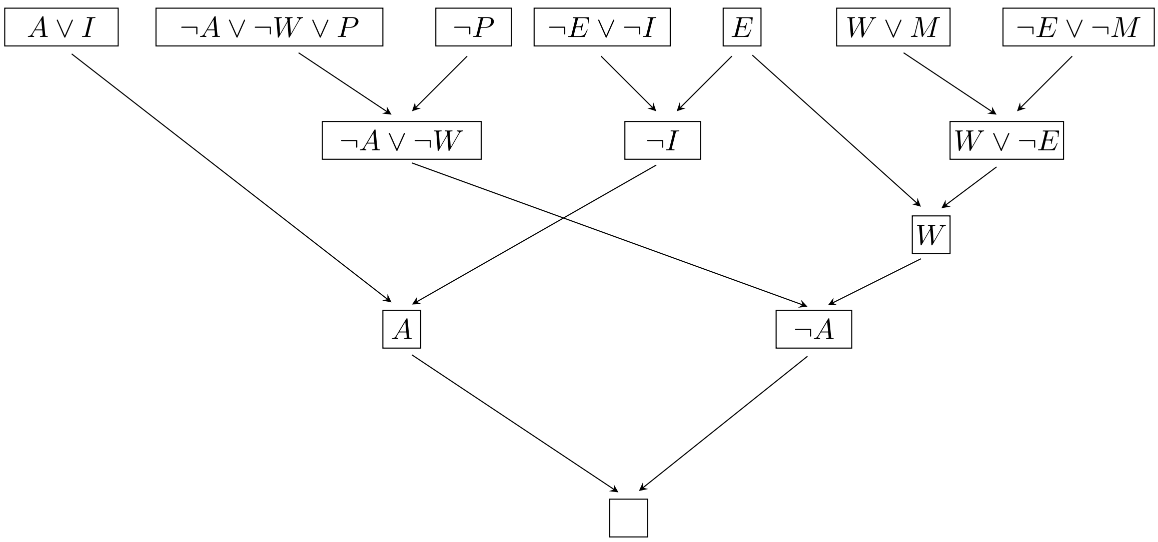

Figure 1: A diagram showing the application of the resolution principle. The clauses on the top are the initial clauses, the ones below them are obtained by combining two pairs of clauses.

Problem 4.5: Completeness and soundness

Recall first the definition of *completeness* and *soundness*:

> **Completeness:** An inference algorithm is **complete** if and only if for every entailed sentence $KB \models \alpha$, the inference algorithm will always be able to derive it.
>
> **Soundness:** An inference algorithm is **sound** if and only if for every sentence it derives, it is guaranteed that the sentence is entailed $KB \models \alpha$.

Problem 4.5.1: Suppose that we have an inference algorithm which will always be able to derive any given sentence (regardless whether it is entailed or not). Would this inference algorithm be complete? Sound?
Solution: This inference algorithm is **complete**, because for every entailed sentence, this algorithm will always be able to derive it. However, this inference algorithm is **unsound**, because it can derive a sentence that is not entailed.

Problem 4.5.2: Suppose now that we have an inference algorithm which will always be unable to derive any sentence (regardless whether it is entailed or not). Would this inference algorithm be complete? Sound?
Solution: This inference algorithm is **incomplete**, because for every entailed sentence, this algorithm will always be unable to derive it. However, this inference algorithm is **sound** since it never derives any sentence.

References

[1] R. Backhouse. *Algorithmic Problem Solving*. Wiley Publishing, 1st edition, 2011.

[2] D. Gries and F. B. Schneider. *A Logical Approach to Discrete Math*. Springer-Verlag New York, Inc., New York, NY, USA, 1993.

[3] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall series in artificial intelligence. Prentice Hall, 2010.